

Method and nucleic acids for the improved treatment of breast cell proliferative disorders

Field of the Invention

In American women, breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer death. In women aged 40-55, breast cancer is the leading cause of death (Greenlee *et al.*, 2000). In 2002, there were 204,000 new cases of breast cancer in the US (data from the American Society of Clinical Oncology) and a comparable number in Europe.

Breast cancer is defined as the uncontrolled proliferation of cells within breasts tissues. Breasts are comprised of 15 to 20 lobes joined together by ducts. Cancer arises most commonly in the duct, but is also found in the lobes with the rarest type of cancer termed inflammatory breast cancer. It will be appreciated by those skilled in the art that there exists a continuing need to improve methods of early detection, classification and treatment of breast cancers. In contrast to the detection of some other common cancers such as cervical and dermal there are inherent difficulties in classifying and detecting breast cancers.

Due to current screening programs and the accessibility of this cancer to self-examination, breast cancer is diagnosed comparatively early: in about 93% of all newly diagnosed cases, the cancer has not yet metastasized, and in 65% of cases, even the lymph nodes are not yet affected.

The first step of any treatment is the assessment of the patient's condition comparative to defined classifications of the disease. However the value of such a system is inherently dependent upon the quality of the classification. Breast cancers are staged according to their size, location and occurrence of metastasis. Methods of treatment include the use of surgery, radiation therapy, chemotherapy and endocrine therapy, which are also used as adjuvant therapies to surgery.

Although the vast majority of early cancers are operable, i.e. the tumor can be completely removed by surgery, about one third of the patients with lymph-node negative diseases and about 50-60% of patients with node-positive disease will develop metastases during follow-up.

Based on this observation, systemic adjuvant treatment has been introduced for both node-positive and node-negative breast cancers. Systemic adjuvant therapy is administered after surgical removal of the tumor, and has been shown to reduce the risk of recurrence significantly (Early Breast Cancer Trialists' Collaborative Group, 1998). Several types of adjuvant treatment are available: endocrine treatment (for hormone receptor positive tumors), different chemotherapy regimens, and novel agents like Herceptin.

The growth of the majority of breast cancers (appr. 70-80%) is dependent on the presence of estrogen. Therefore, one important target for adjuvant therapy is the removal of estrogen (e.g. by ovarian ablation) or the blocking of its actions on the tumor cells (e.g. Tamoxifen). Endocrine treatment is thought to be efficient only in tumors that express hormone receptors (the estrogen receptor, ER, and/or the progesterone receptor, PR). Currently, the vast majority of women with hormone receptor positive breast cancer receive some form of endocrine treatment, independent of their nodal status. The most frequently used drug is Tamoxifen. However, even in hormone receptor positive patients, not all patients benefit from endocrine treatment. Adjuvant endocrine therapy reduces mortality rates by 22% while response rates to endocrine treatment in the advanced setting are 50 to 60% (Jordan *et al.*, 2002, Jordan *et al.*, 1999, Osborne *et al.*, 1998, European Breast Cancer Cooperative Group, 1998).

Since Tamoxifen has relatively few side effects, treatment may be justified even for patients with low likelihood of benefit. However, these patients may require additional, more aggressive adjuvant treatment. This is supported by the fact that, even in earliest and least aggressive tumors, such as node-negative, hormone receptor positive tumours, about 21 % of patients relapse within 10 years after initial diagnosis if they receive Tamoxifen monotherapy as adjuvant treatment (Early Breast Cancer Trialists Collaborative Group. Lancet, 1998).

Several cytotoxic regimens have shown to be effective in reducing the risk of relapse in breast cancer (Mansour *et al.*, 1998). According to current treatment guidelines, most node-positive patients receive adjuvant chemotherapy both in the US and Europe, since the risk of relapse is

considerable. Nevertheless, not all patients do relapse, and there is a proportion of patients who would never have relapsed even without chemotherapy, but who nevertheless receive chemotherapy due to the currently used criteria. In hormone receptor positive patients, chemotherapy is usually given before endocrine treatment, whereas hormone receptor negative patients receive only chemotherapy.

The situation for node-negative patients is particularly complex. In the US, cytotoxic chemotherapy is recommended for node-negative patients, if the tumor is larger than 1 cm. In Europe, chemotherapy is considered for the node-negative cases if one or more risk factors such as tumor size larger than 2 cm, negative hormone receptor status, or tumor grading of three or age <35 is present. In general, there is a tendency to select premenopausal women for additional chemotherapy whereas for postmenopausal women, chemotherapy is often omitted. Compared to endocrine treatment, in particular Tamoxifen, chemotherapy is highly toxic, with short-term side effects such as nausea, vomiting, bone marrow depression, and long-term effects such as cardiotoxicity and an increased risk for secondary cancers.

It is currently not clear which breast cancer patients should be selected for more aggressive therapy, although clinicians agree that there is a need for a subset of patients. The difficulty of selecting the right patients for chemotherapy, and the lack of suitable criteria is also reflected by a recent study which showed that chemotherapy is used much less frequently than recommended, based on data from the New Mexico Tumor registry (Du *et al.*, 2003). This study provides substantial evidence that there is a need for better selection of patients for chemotherapy or other, more aggressive forms of breast cancer therapy.

The levels of observation that have been studied by the methodological developments of recent years in molecular biology, are the genes themselves, the translation of these genes into RNA, and the resulting proteins. The question of which gene is switched on at which point in the course of the development of an individual, and how the activation and inhibition of specific genes in specific cells and tissues are controlled is correlatable to the degree and character of the methylation of the genes or of the genome. In this respect, pathogenic conditions may manifest themselves in a changed methylation pattern of individual genes or of the genome.

DNA methylation plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing since 5-methylcytosine has the same base pairing behaviour as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during PCR amplification.

The currently most frequently used method for analysing DNA for 5-methylcytosine is based upon the specific reaction of bisulphite with cytosine which, upon subsequent alkaline hydrolysis, is converted to uracil which corresponds to thymine in its base pairing behaviour. However, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridisation behaviour, can now be detected as the only remaining cytosine using "normal" molecular biological techniques, for example, by amplification and hybridisation or sequencing. All of these techniques are based on base pairing which can now be fully exploited. In terms of sensitivity, the prior art is defined by a method which encloses the DNA to be analysed in an agarose matrix, thus preventing the diffusion and renaturation of the DNA (bisulphite only reacts with single-stranded DNA), and which replaces all precipitation and purification steps with fast dialysis (Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* 1996 Dec 15; 24(24): 5064-6). Using this method, it is possible to analyse individual cells, which illustrates the potential of the method. However, currently only individual regions of a length of up to approximately 3000 base pairs are analysed, a global analysis of cells for thousands of possible methylation events is not possible. However, this method cannot reliably analyse very small fragments from small sample quantities either. These are lost through the matrix in spite of the diffusion protection.

An overview of the further known methods of detecting 5-methylcytosine may be gathered from the following review article: Rein, T., DePamphilis, M. L., Zorbas, H., *Nucleic Acids Res.* 1998, 26, 2255.

To date, barring few exceptions (e.g., Zeschnigk M, Lich C, Buiting K, Doerfler W, and Horsthemke B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur J Hum Genet.* 1997

Mar-Apr;5(2):94-8) the bisulphite technique is only used in research. Always, however, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment and either completely sequenced (Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet.* 1997 Nov;17(3):275-6) or individual cytosine positions are detected by a primer extension reaction (Gonzalzo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 1997 Jun 15;25(12):2529-31, WO 95/00669) or by enzymatic digestion (Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 1997 Jun 15;25(12):2532-4). In addition, detection by hybridisation has also been described (Olek et al., WO 99/28498).

Further publications dealing with the use of the bisulfite technique for methylation detection in individual genes are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays.* 1994 Jun;16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K, Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Hum Mol Genet.* 1997 Mar;6(3):387-95; Feil R, Charlton J, Bird AP, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.* 1994 Feb 25;22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio MC, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. *Gene.* 1995 May 19;157(1-2):261-4; WO 97/46705, WO 95/15373, and WO 97/45560.

An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999), published in January 1999, and from the literature cited therein.

Fluorescently labelled probes are often used for the scanning of immobilised DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe is particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridised probes may be carried out, for example via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas M, Hillenkamp F. Laser desorption ionisation of proteins with molecular masses exceeding 10,000 Daltons. Anal Chem. 1988 Oct 15;60(20):2299-301). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionised by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones.

MALDI-TOF spectrometry is excellently suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut I G, Beck S. DNA and Matrix Assisted Laser Desorption Ionization Mass Spectrometry. Current Innovations and Future Trends. 1995, 1; 147-57). The sensitivity to nucleic acids is approximately 100 times worse than to peptides and decreases disproportionally with increasing fragment size. For nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is *considerably less efficient*. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity has not been reduced. The difference in sensitivity can be reduced by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. Phosphorothioate nucleic acids in which the usual phosphates of the backbone are substituted with thiophosphates can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut IG, Beck S. A procedure for selective DNA alkylation and detection by mass spectrometry. Nucleic Acids Res. 1995 Apr 25; 23(8): 1367-73). The coupling of a charge tag to this modified DNA results in an increase in sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities which make the detection of unmodified substrates considerably more difficult.

Genomic DNA is obtained from DNA of cell, tissue or other test samples using standard methods. This standard methodology is found in references such as Sambrook, Fritsch and Maniatis, Molecular Cloning: A Laboratory Manual, CSH Press, 2nd edition, 1989: Isolation of genomic DNA from mammalian cells, Protocol I, p. 9.16 - 9.19. Also the manuals of sev-

eral DNA extraction kits such as the QIAamp DNA mini kit give a good guidance on how to isolate genomic DNA.

Currently several predictive markers are under evaluation. As up to now most patients have received Tamoxifen as endocrine treatment most of the markers have been shown to be associated with response or resistance to Tamoxifen. However, it is generally assumed that there is a large overlap between responders to one or the other endocrine treatment. In fact, ER and PR expression are used to select patients for any endocrine treatment. Among the markers which have been associated with TAM response is bcl-2. High bcl-2 levels showed promising correlation to TAM therapy response in patients with metastatic disease and prolonged survival and added valuable information to an ER negative patient subgroup (J Clin Oncology, 1997, 15 5: 1916-1922; Endocrine, 2000, 13(1):1-10). There is conflicting evidence regarding the independent predictive value of c-erbB2 (Her2/neu) overexpression in patients with advanced breast cancer that require further evaluation and verification (British J of Cancer, 1999, 79 (7/8):1220-1226; J Natl Cancer Inst, 1998, 90 (21): 1601-1608).

Other predictive markers include SRC-1 (steroid receptor coactivator-1), CGA gene over expression, cell kinetics and S phase fraction assays (Breast Cancer Res and Treat, 1998, 48:87-92; Oncogene, 2001, 20:6955-6959). Recently, uPA (Urokinase-type plasminogen activator) and PAI-1 (Plasminogen activator inhibitor type 1) together showed to be useful to define a subgroup of patients who have worse prognosis and who would benefit from adjuvant systemic therapy (J Clinical Oncology, 2002, 20 n° 4). However, all of these markers need further evaluations in prospective trials as none of them is yet a validated marker of response.

A number of cancer-associated genes have been shown to be inactivated by hypermethylation of CpG islands during breast tumorigenesis. Decreased expression of the calcium binding protein S100A2 (Accession number NM_005978) has been associated with the development of breast cancers. Hypermethylation of the promoter region of this gene has been observed in neoplastic cells thus providing evidence that S100A2 repression in tumour cells is mediated by site-specific methylation.

The gene SYK (Accession number NM_003177) encodes a protein tyrosine kinase, Syk (spleen tyrosine kinase), that is highly expressed in hematopoietic cells. Syk is expressed in normal breast ductal epithelial cells but not in a subset of invasive breast carcinoma. Also, the

loss of Syk expression seems to be associated with malignant phenotypes such as increased motility and invasion. The loss of expression occurs at the transcriptional level, and, as indicated by Yuan Y, Mendez R, Sahin A and Dai JL (Hypermethylation leads to silencing of the SYK gene in human breast cancer. *Cancer Res.* 2001 Jul 15;61(14):5558-61.), as a result of DNA hypermethylation.

The TGF- β type 2 receptor (encoded by the TGFBR2 gene, NM_003242) plays a role in trans-membrane signalling pathways via a complex of serine/threonine kinases. Mutations in the gene have been detected in some primary tumours and in several types of tumour-derived cell lines, including breast (Lucke CD, Philpott A, Metcalfe JC, Thompson AM, Hughes-Davies L, Kemp PR, Hesketh R. 'Inhibiting mutations in the transforming growth factor beta type 2 receptor in recurrent human breast cancer.' *Cancer Res.* 2001 Jan 15;61(2):482-5.).

The genes COX7A2L and GRIN2D were both identified as novel estrogen responsive elements by Watanabe et. al. (Isolation of estrogen-responsive genes with a CpG island library. *Molec. Cell. Biol.* 18: 442-449, 1998.) using the CpG-GBS (genomic binding site) method. The gene COX7A2L (Accession number NM_004718) encodes a polypeptide 2-like cytochrome C oxidase subunit VIIA. Northern blot analysis detected an upregulation of COX7A2L after estrogen treatment of a breast cancer cell line. The gene GRIN2D (Accession number NM_000836) encodes the N-methyl-D-aspartate, ionotropic, subunit 2D glutamate receptor, a subunit of the NMDA receptor channels associated with neuronal signalling. Furthermore expression of the cDNA has been observed in an osteosarcoma cell line. The gene VTN (also known as Vitronectin Accession number NM_000638) encodes a 75-kD glycoprotein (also called serum spreading factor or complement S-protein) that promotes attachment and spreading of animal cells in vitro, inhibits cytolysis by the complement C5b-9 complex, and modulates antithrombin III-thrombin action in blood coagulation. Furthermore expression of this gene has been linked to progression and invasiveness of cancer cells.

The gene SFN (also known as Stratifin) encodes a polypeptide of the 14-3-3 family, 14-3-3 sigma. The 14-3-3 family of proteins mediates signal transduction by binding to phosphoserine-containing proteins. Expression of the SFN gene is lost in breast carcinomas, this is likely due to hypermethylation during the early stages of neoplastic transformation (see Umbricht CB, Evron E, Gabrielson E, Ferguson A, Marks J, Sukumar S. Hypermethylation of 14-3-3 sigma (stratifin) is an early event in breast cancer. *Oncogene.* 2001 Jun 7; 20(26):3348-53).

The gene PSAT1 (Accession number NM_021154) is not to be confused with the gene popularly referred to as PxySA (Accession number NM_001648) which encodes prostate specific antigen and whose technically correct name is kallikrein 3 . The gene PSAT1 encodes the protein phosphoserine aminotransferase which is the second step-catalysing enzyme in the serine biosynthesis pathway. Changes in gene expression levels have been monitored by mRNA expression analysis and upregulation of the gene has been identified in colonic carcinoma in a study of 6 samples (Electrophoresis 2002 Jun;23(11):1667-76 mRNA differential display of gene expression in colonic carcinoma.Ojala P, Sundstrom J, Gronroos JM, Virtanen E, Talvinen K, Nevalainen TJ).

The gene stathmin (NM_005563) codes for an oncoprotein 18, also known as stathmin, a conserved cytosolic phosphoprotein that regulates microtubule dynamics. The protein is highly expressed in a variety of human malignancies. In human breast cancers the stathmin gene has shown to be up-regulated in a subset of the tumours.

The gene PRKCD encodes a member of the family of protein kinase c enzymes, and is involved in B cell signaling and in the regulation of growth, apoptosis, and differentiation of a variety of cell types.

Some of these molecules interact in a cascade-like manner. PRKCD activity that targets STMN1 is modulated by SFN binding and SYK phosphorylation. Together this influences tubulin polymerization that is required for cell division.

The gene MSMB (Accession number NM_002443) has been mapped to 10q11.2. It encodes the beta-microseminoprotein (MSP) which is one of the major proteins secreted by the prostate. Furthermore, it may be useful as a diagnostic marker for prostate cancer. Using mRNA analysis low levels of beta-MSP mRNA expression and protein have been linked to progression under endocrine therapy and it has been postulated that it may be indicative of potentially aggressive prostate cancer (see Sakai H, Tsurusaki T, Kanda S, Koji T, Xuan JW, Saito Y 'Prognostic significance of beta-microseminoprotein mRNA expression in prostate cancer.' Prostate. 1999 Mar 1;38(4):278-84.).

The gene TP53 (Accession number NM_000546) encodes the protein p53, one of the most well characterised tumour suppressor proteins. The p53 protein acts as a transcription factor and serves as a key regulator of the cell cycle. Inactivation of this gene through mutation disrupts the cell cycle, which, in turn, assists in tumour formation. Methylation changes associated with this gene have been reported to be significant in breast cancer. Saraswati *et. al.* (Nature 405, 974 - 978 (22 Jun 2000) 'Compromised HOXA5 function can limit p53 expression in human breast tumours' reported that low levels of p53 mRNA in breast tumours was correlated to methylation of the HOXA5 gene. The product of the HOXA5 gene binds to the promoter region of the p53 and mediates expression of the gene. Methylation of the promoter region of the p53 gene itself has been reported (Kang JH, Kim SJ, Noh DY, Park IA, Choe KJ, Yoo OJ, Kang HS. 'Methylation in the p53 promoter is a supplementary route to breast carcinogenesis: correlation between CpG methylation in the p53 promoter and the mutation of the p53 gene in the progression from ductal carcinoma in situ to invasive ductal carcinoma.' Lab Invest. 2001 Apr;81(4):573-9.). It was therein demonstrated that CpG methylation in the p53 promoter region is found in breast cancer and it was hypothesised that methylation in the p53 promoter region could be an alternative pathway to neoplastic progression in breast tumours. It has been observed that treatment with Tamoxifen decreases the level of expression of the p53 gene (Farczadi E, Kaszas I, Baki M, Szende B. 'Changes in apoptosis, mitosis, Her-2, p53 and Bcl2 expression in breast carcinomas after short-term tamoxifen treatment.' Neoplasma. 2002;49(2):101-3.)

The gene CYP2D6 (Accession number: NM_000106) is a member of the human cytochrome P450 (CYP) superfamily. Many members of this family are involved in drug metabolism (see for example Curr Drug Metab. 2002 Jun;3(3):289-309. Rodrigues AD, Rushmore TH.), of these Cytochrome P450 CYP2D6 is one of the most extensively characterised. It is highly polymorphic (more than 70 variations of the gene have been described), and allelic variation can result in both increased and decreased enzymatic activity. The CYP2D6 enzyme catalyses the metabolism of a large number of clinically important drugs including antidepressants, neuroleptics, some antiarrhythmics (Nature 1990 Oct 25;347(6295):773-6 Identification of the primary gene defect at the cytochrome P450 CYP2D locus. Gough AC, Miles JS, Spurr NK, Moss JE, Gaedigk A, Eichelbaum M, Wolf CR.).

The gene PTGS2 (Accession number NM_000963) encodes an inducible isozyme of prostaglandin-endoperoxide synthase (prostaglandin-endoperoxide synthase 2). It is also known as

COX2 (cyclooxygenase 2). Aberrant methylation of this gene has been identified in lung carcinomas (Cancer Epidemiol Biomarkers Prev 2002 Mar;11(3):291-7 Hierarchical clustering of lung cancer cell lines using DNA methylation markers. Virmani AK, Tsou JA, Siegmund KD, Shen LY, Long TI, Laird PW, Gazdar AF, Laird-Offringa IA.).

The gene CGA (Accession number NM_000735) encodes the alpha polypeptide of glycoprotein hormones. Further, it has been identified as an estrogen receptor alpha (ER alpha)-responsive gene and overexpression of the gene has been linked to ER positivity in breast tumours. Bieche *et. al.* examined mRNA levels of said gene in 125 ER alpha-positive postmenopausal breast cancer patients treated with primary surgery followed by adjuvant tamoxifen therapy. Initial results indicated significant links between CGA gene overexpression and Scarff-Bloom-Richardson histopathological grade I+II and progesterone and estrogen receptor positivity, which suggested that CGA is a marker of low tumour aggressiveness ('Identification of CGA as a Novel Estrogen Receptor-responsive Gene in Breast Cancer: An Outstanding Candidate Marker to predict the Response to Endocrine Therapy Cancer Research' 61, 1652-1658, February 15, 2001. Ivan Bièche, Béatrice Parfait, Vivianne Le Doussal, Martine Olivi, Marie-Christine Rio, Rosette Lidereau and Michel Vidaud). Further mRNA expression analysis linked CGA expression levels to Tamoxifen response, it was postulated that when combined with analysis of the marker ERBB2 (a marker of poor response) the gene may be useful as a predictive marker of tamoxifen responsiveness in breast cancer (Oncogene 2001 Oct 18;20(47):6955-9 The CGA gene as new predictor of the response to endocrine therapy in ER alpha-positive postmenopausal breast cancer patients. Bieche I, Parfait B, Nogues C, Andrieu C, Vidaud D, Spyrtos F, Lidereau R, Vidaud M.). The authors provided significant data associating the expression of the gene CGA with Tamoxifen treatment response. However, said analyses have all focused upon the analysis of relative levels of mRNA expression. This is not a methodology that is suitable for a medium or high throughput, nor is it a suitable basis for the development of a clinical assay.

The gene PITX2 (NM_000325) encodes the paired-like homeodomain transcription factor 2 which is known to be expressed during development of anterior structures such as the eye, teeth, and anterior pituitary. Although the expression of this gene is associated with cell differentiation and proliferation it has no heretofore recognised role in carcinogenesis or responsiveness to endocrine treatment. Toyota *et al.*, (2001. Blood. 97:2823-9.) found hypermeth-

ylation of the PITX2 gene in a large proportion of acute myeloid leukemias. Furthermore, this hypermethylation is positively correlated to methylation of the ER gene.

RASSF1A (Ras association domain family 1A) gene is a candidate tumour suppressor gene at 3p21.3. The Ras GTPases are a superfamily of molecular switches that regulate cellular proliferation and apoptosis in response to extra-cellular signals. It is purported that *RASSF1A* is a tumour suppressor gene, and epigenetic alterations of this gene have been observed in a variety of cancers. Methylation of *RASSF1A* has been associated with poor prognosis in primary non-small cell lung cancer (Kim DH, Kim JS, Ji YI, Shim YM, Kim H, Han J, Park J., 'Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer.' Cancer Res. 2003 Jul 1;63(13):3743-6.). It has also been associated with the development of pancreatic cancer (Kuzmin I, Liu L, Dammann R, Geil L, Stanbridge EJ, Wilczynski SP, Lerman MI, Pfeifer GP. 'Inactivation of RAS association domain family 1A gene in cervical carcinomas and the role of human papillomavirus infection.' Cancer Res. 2003 Apr 15;63(8):1888-93.), as well as testicular tumours and prostate carcinoma amongst others. The application of the methylation of this gene as a cancer diagnostic marker has been described in U.S. patent 6,596,488, it does not however describe its application in the selection of appropriate treatments regimens for patients.

Also located within 3p21 is the Dystroglycan precursor gene (Dystrophin-associated glycoprotein 1) (NM_004393). Dystroglycan (DG, also known as DAG1) is an adhesion molecule comprising two subunits namely alpha-DG and beta-DG. The molecule is responsible for crucial interactions between extracellular matrix and cytoplasmatic compartment and it has been hypothesised that as such it may contribute to progression to metastatic disease. Decreased expression of this gene has been associated with correlated with higher tumour grade and stage in colon, prostate and breast tumours.

The onecut-2 transcription factor gene (NM_004852) is located at 18q21.31 is a homeo-domain transcription factor regulator of liver gene expression in adults and during development.

The trefoil factor 1 (TFF1) gene (NM_003225) encodes a member of the trefoil family of proteins. The gene is also known as pS2. They are normally expressed at highest levels in the mucosa of the gastrointestinal tract, however they are often expressed ectopically in primary

tumours of other tissues, including breast. The expression of TFF1 is regulated by estrogen in estrogen-responsive breast cancer cells in culture, its expression is associated with that of the estrogen receptor and TFF1 is a marker of hormone responsiveness in tumours (Schwartz *et al.*, 1991. pS2 expression and response to hormonal therapy in patients with advanced breast cancer. *Cancer Res.* 51:624-8). TFF1 promoter methylation has been observed in nonexpressing gastric carcinoma-derived cell lines and tissues.

TMEFF2 (NM_016192) encodes a transmembrane protein containing an epidermal growth factor (EGF)-like motif and two follistatin domains. It has been shown to be overexpressed in prostate and brain tissues and it has been suggested that this is an androgen-regulated gene exhibiting antiproliferative effects in prostate cancer cells.

Methylation of the gene ESR1 (NM_000125), encoding the estrogen receptor has been linked to several cancer types including lung, oesophageal, brain and colorectal. The estrogen receptor (ESR) is a ligand-activated transcription factor composed of several domains important for hormone binding, DNA binding, and activation of transcription. Furthermore, it is the direct target of the anti-estrogenic compound Tamoxifen. Only tumours expressing estrogen receptor (ER+) can respond on Tamoxifen treatment.

The PCAF (NM_003884) gene encodes the p300/CBP-Associated Factor (PCAF). CBP and p300 are large nuclear proteins that bind to many sequence-specific factors involved in cell growth and/or differentiation. The p300/CBP associated factor displays in vivo binding activity with CBP and p300. The protein has histone acetyl transferase activity with core histones and nucleosome core particles, indicating that it plays a direct role in transcriptional regulation. p300/CBP associated factor also associates with NF-kappa-B p65. This protein has been shown to regulate expression of the gene p53 by acetylation of Lys320 in the C-terminal portion of p53.

The WBP11 (NM_016312) gene encodes a nuclear protein, which co-localises with mRNA splicing factors and intermediate filament-containing perinuclear networks. It contains two proline-rich regions that bind to the WW domain of Npw38, a nuclear protein, and thus this protein is also called Npw38-binding protein NpwBP.

The TBC1 domain family, member 3 gene (TBC1D3, NM_032258) was discovered originally as an oncogene, also known as PRC17. The gene product contains a GTPase-activating protein (GAP) catalytic core motif and interacts directly with Rab5, stimulating its GTP hydrolysis. TBC1D3 is amplified in 15% of prostate cancers and highly overexpressed in approximately one-half of metastatic prostate tumors (Pei *et al.*, 2002; Cancer Res. 62:5420-4).

The CDK6 gene encodes a cyclin-dependent protein kinase regulating major cell cycle transitions in eukaryotic cells. The cdk6 kinase is associated with cyclins D1, D2, and D3 and can phosphorylate pRB, the product of the retinoblastoma tumor suppressor gene. The activation of cdk6 kinase occurs during mid-G1 (Meyerson and Harlow, 1994; Mol Cell Biol. 14:2077-86).

Description

In the following certain genetic regions are described for whom no genetic nomenclature is presently available. In each case the chromosomal location of the genetic sequence is denoted within parentheses () and the genetic sequence is further described by its sequence according to Table 1.

The present invention provides methods and nucleic acids for the improved treatment planning of patients with cell proliferative disorders of the breast tissues. The aim of the invention is achieved by assessment of one or both of two factors of particular relevance to patient treatment planning. The first factor is the characterisation of the cell proliferative disorder of the breast tissues and/or a metastases thereof in terms of aggressivity, the second factor being the prediction of disease free survival and/or response of a subject with said disorder to a therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion. Said treatments include, but are not limited to estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.

The prediction of response to a therapeutic regimen comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion (a current treatment of choice as side effects are limited) further enables the physi-

cian to determine if additional treatments will be required in addition to or instead of this treatment. Treatments which may be used in addition to or instead of said treatment include, but are not limited to chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibodies and molecularly targeted drugs.

Characterisation of a breast cancer in terms of its predicted aggressiveness enables the physician to make an informed decision as to a therapeutic regimen with appropriate risk and benefit trade offs to the patient. Aggressiveness is taken to mean one or more of decreased patient survival or disease- or relapse-free survival, increased tumor-related complications and faster progression of tumor or metastases. According to the aggressiveness of the disease an appropriate treatment or treatments may be selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibody treatments, treatments involving molecularly targeted drugs, estrogen receptor modulator treatments, estrogen receptor down-regulator treatments, aromatase inhibitors treatments, ovarian ablation, treatments providing LHRH analogues or other centrally acting drugs influencing estrogen production. Wherein a cancer is characterised as 'aggressive' it is particularly preferred that a treatment such as, but not limited to, chemotherapy is provided in addition to or instead of an endocrine targeting therapy.

Using the methods and nucleic acids described herein, statistically significant models of patient disease free survival and/or responsiveness to treatment and/or disease progression can be developed and utilised to assist patients and clinicians in determining suitable treatment options to be included in the therapeutic regimen. In one aspect the described method is to be used to assess the utility of therapeutic regimens comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion as a therapy for patients suffering from a cell proliferative disorder of the breast tissues. In particular this aspect of the method enables the physician to determine which treatments may be used in addition to or instead of said treatment. In a further aspect the described method enables the characterisation of the cell proliferative disorder in terms of aggressiveness, thereby enabling the physician to recommend suitable treatments. Thus, the present invention will be seen to reduce the problems associated with present breast cell proliferative disorder treatment response prediction methods.

Using the methods and nucleic acids as described herein, patient responsiveness can be evaluated before or during treatment for a cell proliferative disorder of the breast tissues, in order to provide critical information to the patient and clinician as to the likely progression of the disease. It will be appreciated, therefore, that the methods and nucleic acids exemplified herein can serve to improve a patient's quality of life and odds of treatment success by allowing both patient and clinician a more accurate assessment of the patient's treatment options.

The method according to the definition may be used for the improved treatment of all breast cell proliferative disorder patients, both pre and post menopausal and independent of their node or estrogen receptor status. However, it is particularly preferred that said patients are node-negative and estrogen receptor positive.

The aim of the invention is most preferably achieved by means of the analysis of the methylation patterns of one or a combination of genes taken from the group taken from the group EGR4, APC, CDKN2A, CSPG2, ERBB2, STMN1, STK11, CA9, PAX6, SFN, S100A2, TFF1, TGFBR2, TP53, TP73, PLAU, TMEFF2, ESR1, SYK, HSPB1, RASSF1, TES, GRIN2D, PSAT1, CGA, CYP2D6, COX7A2L, ESR2, PLAU, VTN, SULT1A1, PCAF, PRKCD, ONECUT2, BCL6, WBP11, (MX1)MX1, APP, ORC4L, NETO1, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, SEQ ID NO: 48, ABCA8, SEQ ID NO: 50, SEQ ID NO: 51, MARK2, ELK1, Q8WUT3, CGB, BSG, BCKDK, SOX8, DAG1, SEMA4B, and ESR1 (exon8) (see Table 1) and/or their regulatory regions.

The invention is characterised in that the nucleic acid of one or a combination of genes taken from the group EGR4, APC, CDKN2A, CSPG2, ERBB2, STMN1, STK11, CA9, PAX6, SFN, S100A2, TFF1, TGFBR2, TP53, TP73, PLAU, TMEFF2, ESR1, SYK, HSPB1, RASSF1, TES, GRIN2D, PSAT1, CGA, CYP2D6, COX7A2L, ESR2, PLAU, VTN, SULT1A1, PCAF, PRKCD, ONECUT2, BCL6, WBP11, (MX1)MX1, APP, ORC4L, NETO1, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, SEQ ID NO: 48, ABCA8, SEQ ID NO: 50, SEQ ID NO: 51, MARK2, ELK1, Q8WUT3, CGB, BSG, BCKDK, SOX8, DAG1, SEMA4B, and ESR1 (exon8) are contacted with a reagent or series of reagents capable of distinguishing between methylated and non methylated CpG dinucleotides within the genomic sequence of interest.

The present invention makes available a method for the improved treatment and monitoring of breast cell proliferative disorders, by enabling the accurate prediction of a patient's disease free survival and/or response to treatment with a therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production, or secretion.

In a particularly preferred embodiment, the method according to the invention enables the differentiation between patients who have a high probability of response to said therapy and those who have a low probability of response to said therapy or a methylation characteristic predicted disease free survival time, in addition to the characterisation of tumors in terms of aggressiveness.

The method according to the invention may be used for the analysis of a wide variety of cell proliferative disorders of the breast tissues including, but not limited to, ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

The method according to the invention is particularly suited to the prediction of response to the aforementioned therapy in two treatment settings. In one embodiment, the method is applied to patients who receive endocrine pathway targeting treatment as secondary treatment to an initial non chemotherapeutical therapy, e.g. surgery (hereinafter referred to as the adjuvant setting) as illustrated in Figure 1. Such a treatment is often prescribed to patients suffering from Stage 1 to 3 breast carcinomas. In this embodiment patients disease free survival times are predicted according to their by detecting patients with worse disease free survival times the physician may choose to recommend the patient for further treatment, instead of or in addition to the endocrine targeting therapy(s), in particular but not limited to, chemotherapy.

In a further preferred embodiment said method is applied to patients suffering from a relapse of breast cancer following treatment by a primary means (preferably surgery) followed by a disease free period, and wherein the endocrine pathway targeting treatment has been prescribed in response to a detection of a relapse of the carcinoma. Such a treatment is often prescribed to patients suffering from later stage carcinomas, particularly wherein metastasis has

occurred. Therefore this clinical setting shall also hereinafter be referred to as the 'metastatic setting'. In this embodiment responders are those who enter partial or complete remission i.e. subjects whose cancer recedes to undetectable levels as opposed to those whose diseases further metastasise or remain above detectable levels. By detecting patients whose cancers are likely to metastasise the physician may choose to recommend the patient for further treatment, instead of or in addition to the endocrine targetting therapy(s), in particular but not limited to, chemotherapy.

This methodology presents further improvements over the state of the art in that the method may be applied to any subject, independent of the estrogen and/or progesterone receptor status. Therefore in a preferred embodiment, the subject is not required to have been tested for estrogen or progesterone receptor status.

The object of the invention is achieved by means of the analysis of the methylation patterns of one or more of the genes EGR4, APC, CDKN2A, CSPG2, ERBB2, STMN1, STK11, CA9, PAX6, SFN, S100A2, TFF1, TGFBR2, TP53, TP73, PLAU, TMEFF2, ESR1, SYK, HSPB1, RASSF1, TES, GRIN2D, PSAT1, CGA, CYP2D6, COX7A2L, ESR2, PLAU, VTN, SULT1A1, PCAF, PRKCD, ONECUT2, BCL6, WBP11, (MX1)MX1, APP, ORC4L, NETO1, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, SEQ ID NO: 48, ABCA8, SEQ ID NO: 50, SEQ ID NO: 51, MARK2, ELK1, Q8WUT3, CGB, BSG, BCKDK, SOX8, DAG1, SEMA4B, ESR1 (exon8) and/or their regulatory regions. In a particularly preferred embodiment the sequences of said genes comprise SEQ ID NOs: 1-61 and sequences complementary thereto.

The object of the invention may also be achieved by analysing the methylation patterns of one or more genes taken from the following subsets of said aforementioned group of genes. In one embodiment the object of the invention is the prediction of disease free survival and/or probability of response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion. This is achieved by analysis of the methylation patterns of one or more genes taken from the group consisting ERBB2, STMN1, TFF1, TMEFF2, ESR1, HSPB1, PITX2, COX7A2L, PLAU, VTN, PCAF, ONECUT2, BCL6, WBP11, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, ABCA8 and SEQ ID NO: 51 and wherein it is further preferred that the sequence of said genes comprise SEQ ID NOs: 5, 6, 12, 17, 18, 20, 23, 28, 16, 31, 33, 35, 36, 37, 43, 44, 46, 47, 49 and 51, respectively, according to Table 1. It is preferred that said gene is PITX2.

It is preferred that the object of the invention is achieved by analysing the methylation patterns of a plurality of genes, hereinafter also referred to as a gene panel. It is further preferred that said plurality is between two and four genes. PITX2 In one embodiment said gene panel consists of PITX2, TBC1D3 and CDK6. It is particularly preferred that said gene panel of genes is selected from the group consisting TFF1 and PLAU; TFF1 and PLAU and PITX2; PITX2 and TFF1; PITX2 and PLAU. Further preferred is the gene panel of TFF1 and PITX2 for the prediction of disease free survival or metastasis in treated patients.

In a further embodiment the object of the invention is the characterisation of the tumor in terms of aggressiveness. This is achieved by analysis of the methylation patterns of one or more genes taken from the group consisting APC, CSPG2, ERBB2, STK11, S100A2, TFF1, TGFBR2, TP53, TMEFF2, SYK, HSPB1, RASSF1, PSAT1, CGA, ESR2, ONECUT2, WBP11, CYP2D6, CDK6, ELK1, CGB and DAG1, and wherein it is further preferred that the sequence of said genes comprise SEQ ID NOs: 2, 4, 5, 7, 11, 12, 13, 14, 17, 19, 20, 21, 25, 26, 29, 35, 37, 45, 46, 53, 55 and 59, respectively, according to Table 1.

In a preferred embodiment said method is achieved by contacting said nucleic acid sequences in a biological sample obtained from a subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non methylated CpG dinucleotides within the target nucleic acid.

In a preferred embodiment, the method comprises the following steps: Preferably, said method comprises the following steps: In the *first step*, a sample of the tissue to be analysed is obtained. The source may be any suitable source, such as cell lines, histological slides, paraffin embedded tissues, biopsies, tissue embedded in paraffin, bodily fluids, urine, blood and all possible combinations thereof. In a particularly preferred embodiment of the method said source is bodily fluids urine, or blood. The DNA is then isolated from the sample. Extraction may be by means that are standard to one skilled in the art, including the use of commercially available kits, detergent lysates, sonification and vortexing with glass beads. Briefly, wherein the DNA of interest is encapsulated by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a

variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the *second step* of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'pretreatment' herein.

The above-described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis that results in a conversion of non-methylated cytosine nucleobases to uracil or to another base that is dissimilar to cytosine in terms of base pairing behavior.

In the *third step* of the method, fragments of the pretreated DNA are amplified, using sets of primer oligonucleotides according to the present invention, and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NO 206 to 449 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising one or more of SEQ ID NO 1 to 61 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer that hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9

nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 206-449 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

Wherein the method is for the prediction of probability of disease free survival and/or response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion it is particularly preferred that said nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 70, 71, 192, 193, 72, 73, 194, 195, 84, 85, 206, 207, 94, 95, 216, 217, 96, 97, 218, 219, 100, 101, 222, 223, 106, 107, 228, 229, 116, 117, 238, 239, 92, 93, 214, 215, 122, 123, 244, 245, 126, 127, 248, 249, 130, 131, 252, 253, 132, 133, 254, 255, 134, 135, 256, 257, 146, 147, 268, 269, 148, 149, 270, 271, 152, 153, 274, 275, 154, 155, 276, 277, 158, 159, 280, 281, 162, 163, 284 and 285 said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Wherein the method is for the the characterisation of the breast cell proliferative disorder in terms of aggressiveness it is particularly preferred that said nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 64, 65, 186, 187, 68, 69, 190, 191, 70, 71, 192, 193, 74, 75, 196, 197, 82, 83, 204, 205, 84, 85, 206, 207, 86, 87, 208, 209, 88, 89, 210, 211, 94, 95, 216, 217, 98, 99, 220, 221, 100, 101, 222, 223, 102, 103, 224, 225, 110, 111, 232, 233, 112, 113, 234, 235, 118, 119, 240, 241, 130, 131, 252, 253, 134, 135, 256, 257, 150, 151, 272, 273, 152, 153, 274, 275, 166, 167, 288, 289, 170, 171, 292, 293, 178, 179, 300, 301, 148, 149, 270, 271, 150, 151, 272, 273, 152, 153, 274, 275, 154, 155, 276, 277, 156, 157, 278, 279, 158, 159, 280, 281, 160, 161, 282, 283, 162, 163, 284, 285, 164, 165, 286, 287, 166, 167, 288, 289, 168, 169, 290, 291, 170, 171, 292, 293, 172, 173, 294, 295, 174, 175, 296, 297, 176, 177, 298, 299, 178, 179, 300, 301, 180, 181, 302, 303, 182, 183, 304 and 305, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

A further preferred embodiment of the method comprises the use of *blocker* oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is sup-

pressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methylated nucleic acids is desired.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (*e.g.*, with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker - a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase. Preferably, therefore, the base sequence of said *blocking oligonucleotides* is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 206-449, and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG, TpG or CpA dinucleotide.

Wherein the method is for the prediction of probability of disease free survival and/or response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion it is particularly preferred that said nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 70, 71, 192, 193, 72, 73, 194, 195, 84, 85, 206, 207, 94, 95, 216, 217, 96, 97, 218, 219, 100, 101, 222, 223, 106, 107, 228, 229, 116, 117, 238, 239, 92, 93, 214, 215, 122, 123, 244, 245, 126, 127, 248, 249, 130, 131, 252, 253, 132, 133, 254, 255, 134, 135, 256, 257, 146, 147, 268, 269, 148, 149, 270, 271, 152, 153, 274, 275, 154, 155, 276, 277, 158, 159, 280, 281, 162, 163, 284 and 285, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Wherein the method is for the the characterisation of the breast cell proliferative disorder in terms of aggressiveness it is particularly preferred that said nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 64, 65, 186, 187, 68, 69, 190, 191, 70, 71, 192, 193, 74, 75, 196, 197, 82, 83, 204, 205, 84, 85, 206, 207, 86, 87, 208, 209, 88, 89, 210, 211, 94, 95, 216, 217, 98, 99, 220, 221, 100, 101, 222, 223, 102, 103, 224, 225, 110, 111, 232, 233, 112, 113, 234, 235, 118, 119, 240, 241, 130, 131, 252, 253, 134, 135, 256, 257, 150, 151, 272, 273, 152, 153, 274, 275, 166, 167, 288, 289, 170, 171, 292, 293, 178, 179, 300, 301, 148, 149, 270, 271, 150, 151, 272, 273, 152, 153, 274, 275, 154, 155, 276, 277, 156, 157, 278, 279, 158, 159, 280, 281, 160, 161, 282, 283, 162, 163, 284, 285, 164, 165, 286, 287, 166, 167, 288, 289, 168, 169, 290, 291, 170, 171, 292, 293, 172, 173, 294, 295, 174, 175, 296, 297, 176, 177, 298, 299, 178, 179, 300, 301, 180, 181, 302, 303, 182, 183, 304 and 305, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass that can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, *e.g.*, matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas and Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut and Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut and Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In the *fourth step* of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

Amplificates obtained by means of both standard and methylation specific PCR may be further analyzed by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplificates synthesised in *step three* are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO 1 to 61, and the equivalent positions within SEQ ID NO 206-449 (according to Table 1). Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplificates are then removed. The hybridized amplificates are then detected. In this context, it is preferred that labels attached to the amplificates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™ PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Ap-

plied Biosystems, Foster City, California). The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred embodiments, is designed to hybridize to a GpC-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent “reporter moiety” and a “quencher moiety” covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight™ assay. Variations on the TaqMan™ detection methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

A further suitable method for the use of probe oligonucleotides for the assessment of methylation by analysis of bisulfite treated nucleic acids. In a further preferred embodiment of the method, the *fifth step* of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In yet a further embodiment of the method, the *fifth step* of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the *third step* of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

In one preferred embodiment of the method the nucleic acids according to SEQ ID NO 1 to 61, are isolated and treated according to the first three steps of the method outlined above, namely:

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
 - b) extracting or otherwise isolating the genomic DNA; and
 - c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;
- and wherein the subsequent amplification of d) is carried out in a methylation specific manner, namely by use of methylation specific primers or *blocking oligonucleotides*, and further

wherein the detection of the amplicates is carried out by means of a real-time detection probes, as described above.

Wherein the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 206-449, and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

Wherein the method is for the prediction of disease free survival and/or probability of response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion it is particularly preferred that said *blocking oligonucleotide* nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of one of SEQ ID NO 70, 71, 192, 193, 72, 73, 194, 195, 84, 85, 206, 207, 94, 95, 216, 217, 96, 97, 218, 219, 100, 101, 222, 223, 106, 107, 228, 229, 116, 117, 238, 239, 92, 93, 214, 215, 122, 123, 244, 245, 126, 127, 248, 249, 130, 131, 252, 253, 132, 133, 254, 255, 134, 135, 256, 257, 146, 147, 268, 269, 148, 149, 270, 271, 152, 153, 274, 275, 154, 155, 276, 277, 158, 159, 280, 281, 162, 163, 284 and 285, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Wherein the method is for the characterisation of the breast cell proliferative disorder in terms of aggressiveness it is particularly preferred that said *blocking oligonucleotide* nucleotide sequence(s) hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 64, 65, 186, 187, 68, 69, 190, 191, 70, 71, 192, 193, 74, 75, 196, 197, 82, 83, 204, 205, 84, 85, 206, 207, 86, 87, 208, 209, 88, 89, 210, 211, 94, 95, 216, 217, 98, 99, 220, 221, 100, 101, 222, 223, 102, 103, 224, 225, 110, 111, 232, 233, 112, 113, 234, 235, 118, 119, 240, 241, 130, 131, 252, 253, 134, 135, 256, 257, 150, 151, 272, 273, 152, 153, 274, 275, 166, 167, 288, 289, 170, 171, 292, 293, 178, 179, 300, 301, 148, 149, 270, 271, 150, 151, 272, 273, 152, 153, 274, 275, 154, 155, 276, 277, 156, 157, 278, 279, 158, 159, 280, 281, 160, 161, 282, 283, 162, 163, 284, 285, 164, 165, 286, 287, 166, 167, 288, 289, 168, 169, 290, 291, 170, 171, 292, 293, 172, 173, 294, 295, 174, 175, 296, 297, 176, 177, 298, 299, 178, 179, 300, 301, 180, 181, 302, 303, 182, 183, 304 and 305, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Step e) of the method, namely the detection of the specific amplificates indicative of the methylation status of one or more CpG positions according to SEQ ID NO 1 to 61 is carried out by means of real-time detection methods as described above.

In an alternative most preferred embodiment of the method the subsequent amplification of d) is carried out in the presence of *blocking oligonucleotides*, as described above. Said *blocking oligonucleotides* comprising a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 206-449 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide. Step e) of the method, namely the detection of the specific amplificates indicative of the methylation status of one or more CpG positions according to SEQ ID NO 206-449 is carried out by means of real-time detection methods as described above.

In a further preferred embodiment of the method the nucleic acids according to SEQ ID NO 1 to 61 are isolated and treated according to the first three steps of the method outlined above, namely:

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
- b) extracting or otherwise isolating the genomic DNA;
- c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties; and wherein
- d) amplifying subsequent to treatment in c) is carried out in a methylation specific manner, namely by use of methylation specific primers or *blocking oligonucleotides*, and further wherein
- e) detecting of the amplificates is carried out by means of a real-time detection probes, as described above.

Wherein the subsequent amplification of c) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO 206-449 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide. Wherein the method is for the prediction of disease free survival and/or probability of response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism,

production or secretion it is particularly preferred that said methylation specific primers hybridize to a pretreated nucleic acid sequence according to one of one of SEQ ID NO 70, 71, 192, 193, 72, 73, 194, 195, 84, 85, 206, 207, 94, 95, 216, 217, 96, 97, 218, 219, 100, 101, 222, 223, 106, 107, 228, 229, 116, 117, 238, 239, 92, 93, 214, 215, 122, 123, 244, 245, 126, 127, 248, 249, 130, 131, 252, 253, 132, 133, 254, 255, 134, 135, 256, 257, 146, 147, 268, 269, 148, 149, 270, 271, 152, 153, 274, 275, 154, 155, 276, 277, 158, 159, 280, 281, 162, 163, 284 and 285, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Wherein the method is for the characterisation of the breast cell proliferative disorder in terms of aggressiveness it is particularly preferred that said methylation specific primers hybridize to a pretreated nucleic acid sequence according to one of SEQ ID NO 64, 65, 186, 187, 68, 69, 190, 191, 70, 71, 192, 193, 74, 75, 196, 197, 82, 83, 204, 205, 84, 85, 206, 207, 86, 87, 208, 209, 88, 89, 210, 211, 94, 95, 216, 217, 98, 99, 220, 221, 100, 101, 222, 223, 102, 103, 224, 225, 110, 111, 232, 233, 112, 113, 234, 235, 118, 119, 240, 241, 130, 131, 252, 253, 134, 135, 256, 257, 150, 151, 272, 273, 152, 153, 274, 275, 166, 167, 288, 289, 170, 171, 292, 293, 178, 179, 300, 301, 148, 149, 270, 271, 150, 151, 272, 273, 152, 153, 274, 275, 154, 155, 276, 277, 156, 157, 278, 279, 158, 159, 280, 281, 160, 161, 282, 283, 162, 163, 284, 285, 164, 165, 286, 287, 166, 167, 288, 289, 168, 169, 290, 291, 170, 171, 292, 293, 172, 173, 294, 295, 174, 175, 296, 297, 176, 177, 298, 299, 178, 179, 300, 301, 180, 181, 302, 303, 182, 183, 304 and 305, said contiguous nucleotides comprising at least one CpG, TpG or CpA dinucleotide sequence.

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NO 1 to 61), and complements thereof) without the need for pretreatment.

Wherein the method is for the prediction of disease free survival and/or probability of response to a treatment which targets the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion it is particularly preferred that said genomic sequences are selected from SEQ ID NO 5, 6, 12, 17, 18, 20, 23, 28, 16, 31, 33, 35, 36, 37, 43, 44, 46, 47, 49 and 51.

Wherein the method is for the characterisation of the breast cell proliferative disorder in terms of aggressiveness it is particularly preferred that said genomic sequences are selected from SEQ ID NO 2, 4, 5, 7, 11, 12, 13, 14, 17, 19, 20, 21, 25, 26, 29, 35, 37, 45, 46, 53, 55 and 59.

In the *first step* of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, paraffin embedded tissues, body fluids, or tissue embedded in paraffin. In the *second step*, the genomic DNA is extracted. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, in particular with methylation-sensitive restriction endonucleases.

In the *third step*, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the *fourth step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplificates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionucleotides and mass labels.

In the *fifth step* the amplificates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

When the methylation status of the selected CpG positions have been ascertained patient treatment relevant parameters can be ascertained wherein hypermethylation of the genes is associated with poor prognosis of said subject, aggressive characteristics of said cell proliferative disorder.

erative disorder, poor disease free survival and/or lower probability of response of said subject to said treatment as relative to individuals with hypomethylation.

The term "hypermethylation" refers to the average methylation state corresponding to an *increased* (above average or median) presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a control DNA sample.

The term "hypomethylation" refers to the average methylation state corresponding to a *decreased* (below average or median) presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a control DNA sample.

Kits

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent; a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NO: 1 to 61 and 206-449; oligonucleotides and/or PNA-oligomers; as well as instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight™, HeavyMethyl™, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

Typical reagents (*e.g.*, as might be found in a typical MethyLight®-based kit) for MethyLight® analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimised PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimised PCR buffers and deoxynucleotides; gel ex-

traction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

In order to enable the disclosed method, the invention further provides the modified DNA of one or a combination of genes taken from the group EGR4, APC, CDKN2A, CSPG2, ERBB2, STMN1, STK11, CA9, PAX6, SFN, S100A2, TFF1, TGFBR2, TP53, TP73, PLAU, TMEFF2, ESR1, SYK, HSPB1, RASSF1, TES, GRIN2D, PSAT1, CGA, CYP2D6, COX7A2L, ESR2, PLAU, VTN, SULT1A1, PCAF, PRKCD, ONECUT2, BCL6, WBP11, (MX1)MX1, APP, ORC4L, NETO1, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, SEQ ID NO: 48, ABCA8, SEQ ID NO: 50, SEQ ID NO: 51, MARK2, ELK1, Q8WUT3, CGB, BSG, BCKDK, SOX8, DAG1, SEMA4B and ESR1 (exon8) as well as oligonucleotides and/or PNA-oligomers for detecting cytosine methylations within said genes. The present invention is based on the discovery that genetic and epigenetic parameters and, in particular, the cytosine methylation patterns of said genomic DNAs are particularly suitable for improved treatment and monitoring of breast cell proliferative disorders.

The nucleic acids according to the present invention can be used for the analysis of genetic and/or epigenetic parameters of genomic DNA.

This objective according to the present invention is achieved using a nucleic acid containing a sequence of at least 16 bases in length of the pretreated genomic DNA according to one of SEQ ID NO: 206 to SEQ ID NO: 449 and sequences complementary thereto.

The modified nucleic acids could heretofore not be connected with the improved treatment of breast cell proliferative disorders by prediction of disease free survival and/or probability of response to treatment and/or characterisation of the disease in terms of aggressiveness.

The object of the present invention is further achieved by an oligonucleotide or oligomer for the analysis of pretreated DNA, for detecting the genomic cytosine methylation state, said oligonucleotide containing at least one base sequence having a length of at least 10 nucleotides which hybridises to a pretreated genomic DNA according to SEQ ID NO: 206 to SEQ ID NO: 449 . The oligomer probes according to the present invention constitute important and effective tools which, for the first time, make it possible to ascertain specific genetic and epigenetic parameters during the analysis of biological samples for features associated with a patient's disease free survival and/or response to endocrine treatment. Said oligonucleotides allow the improved treatment and monitoring of breast cell proliferative disorders. The base sequence of the oligomers preferably contains at least one CpG or TpG dinucleotide. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Particularly preferred are oligonucleotides according to the present invention in which the cytosine of the CpG dinucleotide is within the middle third of said oligonucleotide e.g. the 5th - 9th nucleotide from the 5'-end of a 13-mer oligonucleotide; or in the case of PNA-oligomers, it is preferred for the cytosine of the CpG dinucleotide to be the 4th - 6th nucleotide from the 5'-end of the 9-mer.

The oligomers according to the present invention are normally used in so called "sets" which contain upto two oligomers and up to one oligomer for each of the CpG dinucleotides within SEQ ID NO: 206 to SEQ ID NO: 449 .

In the case of the sets of oligonucleotides according to the present invention, it is preferred that at least one oligonucleotide is bound to a solid phase. It is further preferred that all the oligonucleotides of one set are bound to a solid phase.

The present invention further relates to a set of at least 2 n (oligonucleotides and/or PNA-oligomers) used for detecting the cytosine methylation state of genomic DNA, by analysis of said sequence or treated versions of said sequence (of the genes EGR4, APC, CDKN2A, CSPG2, ERBB2, STMN1, STK11, CA9, PAX6, SFN, S100A2, TFF1, TGFBR2, TP53, TP73, PLAU, TMEFF2, ESR1, SYK, HSPB1, RASSF1, TES, GRIN2D, PSAT1, CGA, CYP2D6, COX7A2L, ESR2, PLAU, VTN, SULT1A1, PCAF, PRKCD, ONECUT2, BCL6, WBP11, (MX1)MX1, APP, ORC4L, NETO1, TBC1D3, GRB7, CDK6, SEQ ID NO: 47, SEQ ID NO: 48, ABCA8, SEQ ID NO: 50, SEQ ID NO: 51, MARK2, ELK1, Q8WUT3,

CGB, BSG, BCKDK, SOX8, DAG1, SEMA4B, ESR1 (exon8) as detailed in the sequence listing and Table 1) and sequences complementary thereto). These probes enable improved treatment and monitoring of breast cell proliferative disorders.

It will be obvious to one skilled in the art that the method according to the invention will be improved and supplemented by the incorporation of markers and clinical indicators known in the state of the art and currently used as predictive of the outcome of therapies which target endocrine or endocrine associated pathways. More preferably said markers include node status, age, menopausal status, grade, estrogen and progesterone receptors.

The genes that form the basis of the present invention may be used to form a "gene panel", i.e. a collection comprising the particular genetic sequences of the present invention and/or their respective informative methylation sites. The formation of gene panels allows for a quick and specific analysis of specific aspects of breast cancer treatment. The gene panel(s) as described and employed in this invention can be used with surprisingly high efficiency for the treatment of breast cell proliferative disorders by prediction of the outcome of treatment with a therapy comprising one or more drugs which target the estrogen receptor pathway or are involved in estrogen metabolism, production, or secretion. The analysis of each gene of the panel contributes to the evaluation of patient responsiveness, however, in a less preferred embodiment the patient evaluation may be achieved by analysis of only a single gene. The analysis of a single member of the 'gene panel' would enable a cheap but less accurate means of evaluating patient responsiveness, the analysis of multiple members of the panel would provide a rather more expensive means of carrying out the method, but with a higher accuracy (the technically preferred solution).

The efficiency of the method according to the invention is improved when applied to patients who have not been treated with chemotherapy. Accordingly, it is a particularly preferred embodiment of the method wherein the method is used for the assessment of subjects who have not undergone chemotherapy.

According to the present invention, it is preferred that an arrangement of different oligonucleotides and/or PNA-oligomers (a so-called "array") made available by the present invention is present in a manner that it is likewise bound to a solid phase. This array of different oligonucleotide- and/or PNA-oligomer sequences can be characterised in that it is arranged on the

solid phase in the form of a rectangular or hexagonal lattice. The solid phase surface is preferably composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. However, nitrocellulose as well as plastics such as nylon which can exist in the form of pellets or also as resin matrices are suitable alternatives.

Therefore, a further subject matter of the present invention is a method for manufacturing an array fixed to a carrier material for the improved treatment and monitoring of breast cell proliferative disorders. In said method at least one oligomer according to the present invention is coupled to a solid phase. Methods for manufacturing such arrays are known, for example, from US Patent 5,744,305 by means of solid-phase chemistry and photolabile protecting groups.

A further subject matter of the present invention relates to a DNA chip for the improved treatment and monitoring of breast cell proliferative disorders. The DNA chip contains at least one nucleic acid according to the present invention. DNA chips are known, for example, in US Patent 5,837,832.

The oligomers according to the present invention or arrays thereof as well as a kit according to the present invention are intended to be used for the improved treatment and monitoring of breast cell proliferative disorders. According to the present invention, the method is preferably used for the analysis of important genetic and/or epigenetic parameters within genomic DNA, in particular for use in improved treatment and monitoring of breast cell proliferative disorders.

The methods according to the present invention are used, for improved treatment and monitoring of breast cell proliferative disorder by enabling more informed therapeutic regimens.

The present invention moreover relates to the diagnosis and/or prognosis of events which are disadvantageous or relevant to patients or individuals in which important genetic and/or epigenetic parameters within genomic DNA, said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for the diagnosis and/or prognosis of events which are disadvantageous or relevant to patients or individuals.

In the context of the present invention the term "hybridisation" is to be understood as a bond of an oligonucleotide to a completely complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

In the context of the present invention, "genetic parameters" are mutations and polymorphisms of genomic DNA and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

In the context of the present invention the term "methylation state" is taken to mean the degree of methylation present in a nucleic acid of interest, this may be expressed in absolute or relative terms i.e. as a percentage or other numerical value or by comparison to another tissue and therein described as hypermethylated, hypomethylated or as having significantly similar or identical methylation status.

In the context of the present invention the term "regulatory region" of a gene is taken to mean nucleotide sequences which affect the expression of a gene. Said regulatory regions may be located within, proximal or distal to said gene. Said regulatory regions include but are not limited to constitutive promoters, tissue-specific promoters, developmental-specific promoters, inducible promoters and the like. Promoter regulatory elements may also include certain enhancer sequence elements that control transcriptional or translational efficiency of the gene.

In the context of the present invention the term "chemotherapy" is taken to mean the use of drugs or chemical substances to treat cancer. This definition excludes radiation therapy (treatment with high energy rays or particles), hormone therapy (treatment with hormones or hormone analogues (synthetic substitutes) and surgical treatment.

In the context of the present invention, "epigenetic parameters" are, in particular, cytosine methylations and further modifications of DNA bases of genomic DNA and sequences further required for their regulation. Further epigenetic parameters include, for example, the acetylation of histones which, cannot be directly analysed using the described method but which, in turn, correlates with the DNA methylation.

In the context of the present invention the term "adjuvant treatment" is taken to mean a therapy of a cancer patient immediately following an initial non chemotherapeutical therapy, e.g. surgery. In general, the purpose of an adjuvant therapy is to provide a significantly smaller risk of recurrences compared without the adjuvant therapy.

In the context of the present invention the terms "estrogen receptor positive" and/or "progesterone receptor positive" when used to describe a breast cell proliferative disorder are taken to mean that the proliferating cells express said hormone receptor.

BEST MODE

Characterization of a breast cancer in terms of prognosis and/or treatment outcome enables the physician to make an informed decision as to a therapeutic regimen with appropriate risk and benefit trade off's to the patient.

In the context of the present mode of the invention the terms "estrogen receptor positive" and/or "progesterone receptor positive" when used to describe a breast cell proliferative disorder are taken to mean that the proliferating cells express said hormone receptor.

In the context of the present mode of the invention the term 'aggressiveness' is taken to mean one or more of high likelihood of relapse post surgery; below average or below median patient survival; below average or below median disease free survival; below average or below median relapse-free survival; above average tumor-related complications; fast progression of tumor or metastases. According to the aggressiveness of the disease an appropriate treatment or treatments may be selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibody treatments, treatments involving molecularly targeted drugs, estrogen receptor modulator treatments, estrogen receptor down-regulator treatments, aromatase inhibitors treatments, ovarian ablation, treatments providing LHRH analogues or other centrally acting drugs influencing estrogen production. Wherein a cancer is characterized as 'aggressive' it is particularly preferred that a treatment such as, but not limited to, chemotherapy is provided in addition to or instead of an endocrine targeting therapy. Indicators of tumor aggressiveness standard in the art include but are not limited to, tumor stage, tumor grade, nodal status and survival.

Unless stated otherwise as used herein the term "survival" shall be taken to include all of the following: survival until mortality, also known as overall survival (wherein said mortality may be either irrespective of cause or breast tumor related); "recurrence-free survival" (wherein the term recurrence shall include both localized and distant recurrence) ; metastasis free survival; disease free survival (wherein the term disease shall include breast cancer and diseases associated therewith). The length of said survival may be calculated by reference to a defined start point (e.g. time of diagnosis or start of treatment) and end point (e.g. death, recurrence or metastasis).

As used herein the term "prognostic marker" shall be taken to mean an indicator of the likelihood of progression of the disease, in particular aggressiveness and metastatic potential of a breast tumor.

As used herein the term 'predictive marker' shall be taken to mean an indicator of response to therapy, said response is preferably defined according to patient survival. It is preferably used to define patients with high, low and intermediate length of survival or recurrence after treatment, that is the result of the inherent heterogeneity of the disease process.

As defined herein the term predictive marker may in some situations fall within the remit of a herein described 'prognostic marker', for example, wherein a prognostic marker differentiates between patients with different survival outcomes pursuant to a treatment, said marker is also a predictive marker for said treatment. Therefore, unless otherwise stated the two terms shall not be taken to be mutually exclusive.

As used herein the term 'expression' shall be taken to mean the transcription and translation of a gene, as well as the genetic or the epigenetic modifications of the genomic DNA associated with the marker gene and/or regulatory or promoter regions thereof. Genetic modifications include SNPs, point mutations, deletions, insertions, repeat length, rearrangements and other polymorphisms. The analysis of either the expression levels of protein, or mRNA or the analysis of the patient's individual genetic or epigenetic modification of the marker gene are herein summarized as the analysis of 'expression of the gene.

The level of expression of a gene may be determined by the analysis of any factors associated with or indicative of the level of transcription and translation of a gene including but not lim-

ited to methylation analysis, loss of heterozygosity (hereinafter also referred to as LOH), RNA expression levels and protein expression levels.

Furthermore the activity of the transcribed gene may be affected by genetic variations such as but not limited to genetic modifications (including but not limited to SNPs, point mutations, deletions, insertions, repeat length, rearrangements and other polymorphisms).

The terms "endocrine therapy" or "endocrine treatment" are meant to comprise any therapy, treatment or treatments targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway, which is involved in estrogen metabolism, production or secretion. Said treatments include, but are not limited to estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.

The term "monotherapy" shall be taken to mean the use of a single drug or other therapy.

In the context of the present mode of the invention the term "chemotherapy" is taken to mean the use of pharmaceutical or chemical substances to treat cancer. This definition excludes radiation therapy (treatment with high energy rays or particles), hormone therapy (treatment with hormones or hormone analogues) and surgical treatment.

In the context of the present mode of the invention the term "adjuvant treatment" is taken to mean a therapy of a cancer patient immediately following an initial non chemotherapeutical therapy, e.g. surgery. In general, the purpose of an adjuvant therapy is to decrease the risk of recurrence.

In the context of the present mode of the invention the term "determining a suitable treatment regimen for the subject" is taken to mean the determination of a treatment regimen (i.e. a single therapy or a combination of different therapies that are used for the prevention and/or treatment of the cancer in the patient) for a patient that is started, modified and/or ended based or essentially based or at least partially based on the results of the analysis according to the present mode of the invention. One example is starting an adjuvant endocrine therapy after surgery, another would be to modify the dosage of a particular chemotherapy. The determination can, in addition to the results of the analysis according to the present mode of the inven-

tion, be based on personal characteristics of the subject to be treated. In most cases, the actual determination of the suitable treatment regimen for the subject will be performed by the attending physician or doctor.

In the context of this mode of the invention the terms "obtaining a biological sample" or "obtaining a sample from a subject", shall not be taken to include the active retrieval of a sample from an individual, e.g. the performance of a biopsy. Said terms shall be taken to mean the obtainment of a sample previously isolated from an individual. Said samples may be isolated by any means standard in the art, including but not limited to biopsy, surgical removal, body fluids isolated by means of aspiration. Furthermore said samples may be provided by third parties including but not limited to clinicians, couriers, commercial sample providers and sample collections.

In the context of the present mode of the invention, the term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6 , and (2) having a "GC Content" >0.5 . CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

In the context of the present mode of the invention the term "regulatory region" of a gene is taken to mean nucleotide sequences which affect the expression of a gene. Said regulatory regions may be located within, proximal or distal to said gene. Said regulatory regions include but are not limited to constitutive promoters, tissue-specific promoters, developmental-specific promoters, inducible promoters and the like. Promoter regulatory elements may also include certain enhancer sequence elements that control transcriptional or translational efficiency of the gene.

In the context of the present mode of the invention, the term "methylation" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence.

In the context of the present mode of the invention the term "methylation state" is taken to mean the degree of methylation present in a nucleic acid of interest, this may be expressed in absolute or relative terms i.e. as a percentage or other numerical value or by comparison to

another tissue and therein described as hypermethylated, hypomethylated or as having significantly similar or identical methylation status.

In the context of the present mode of the invention, the term "hemi-methylation" or "hemimethylation" refers to the methylation state of a CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the double stranded CpG methylation site is methylated (*e.g.*, 5'-NNC^MGNN-3' (top strand): 3'-NNGCNN-5' (bottom strand)).

In the context of the present mode of the invention, the term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present mode of the invention, the term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present mode of the invention, the term "microarray" refers broadly to both "DNA microarrays," and "DNA chip(s)," as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as genetic modifications or mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic modifications" or "epigenetic parameters" are modifications of DNA bases of genomic DNA and sequences further required for their regulation, in particular, cytosine methylations thereof. Further epigenetic parameters include, for example, the acetylation of

histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

In the context of the present mode of the invention, the term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

In the context of the present mode of the invention, the term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

In the context of the present mode of the invention, the term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

In the context of the present mode of the invention, the term "MethyLight" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

In the context of the present mode of the invention, the term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a methylation assay comprising methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In the context of the present mode of the invention the term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

In the context of the present mode of the invention the term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong and Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

In the context of the present mode of the invention the term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley and Sons, N.Y.) at Unit 2.10.

"Background DNA" as used herein refers to any nucleic acids which originate from sources other than breast cells.

Using the methods and nucleic acids described herein, statistically significant models of patient relapse, disease free survival, metastasis free survival, overall survival and/or disease progression can be developed and utilized to assist patients and clinicians in determining suitable treatment options to be included in the therapeutic regimen.

In one aspect the method provides a prognostic marker for a cell proliferative disorder of the breast tissues. Preferably this prognosis is provided in terms of an outcome selected from the group consisting of likelihood of relapse; overall patient survival; metastasis free survival; disease free survival or disease progression.

In a further aspect of the invention said marker is used as a predictive marker of outcome of a treatment which targets the estrogen receptor pathway or is involved in estrogen metabolism, production or secretion as a therapy for patients suffering from a cell proliferative disorder of the breast tissues. This aspect of the method enables the physician to determine which treatments may be used in addition to or instead of said endocrine treatment. It is preferred that said additional treatment is a more aggressive therapy such as, but not limited to, chemotherapy. Thus, the present invention will be seen to reduce the problems associated with present breast cell proliferative disorder prognostic, predictive and treatment response prediction methods.

Using the methods and nucleic acids as described herein, patient survival can be evaluated before or during treatment for a cell proliferative disorder of the breast tissues, in order to provide critical information to the patient and clinician as to the likely progression of the disease. It will be appreciated, therefore, that the methods and nucleic acids exemplified herein can serve to improve a patient's quality of life and odds of treatment success by allowing both patient and clinician a more accurate assessment of the patient's treatment options.

The herein disclosed method may be used for the improved treatment of all breast cell proliferative disorder patients, both pre- and post- menopausal and independent of their node or estrogen receptor status. However, it is particularly preferred that said patients are node-negative and estrogen receptor positive.

The present invention makes available a method for the improved treatment of breast cell proliferative disorders, by enabling the improved prediction of a patient's survival, in particular by predicting the likelihood of relapse post-surgery both with or without adjuvant endocrine treatment. Furthermore, the present invention provides a means for the improved prediction of treatment outcome with endocrine therapy, wherein said therapy comprises one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production, or secretion.

The method according to the invention may be used for the analysis of a wide variety of cell proliferative disorders of the breast tissues including, but not limited to, ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid

carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

The method according to the invention may be used to provide a prognosis of breast cell proliferative disorder patients, furthermore said method may be used to provide a prediction of patient survival and/or relapse following treatment by endocrine therapy.

Wherein the herein disclosed markers, methods and nucleic acids are used as prognostic markers it is particularly preferred that said prognosis is defined in terms of patient survival and/or relapse. In this embodiment patients survival times and/or relapse are predicted according to their gene expression or genetic or epigenetic modifications thereof. In this aspect of the invention it is particularly preferred that said patients are tested prior to receiving any adjuvant endocrine treatment.

Wherein the herein disclosed markers, methods and nucleic acids are used as predictive markers it is particularly preferred that the method is applied to predict the outcome of patients who receive endocrine treatment as secondary treatment to an initial non chemotherapeutical therapy, e.g. surgery (hereinafter referred to as the 'adjuvant setting') as illustrated in Figure 1. Such a treatment is often prescribed to patients suffering from Stage 1 to 3 breast carcinomas. It is also preferred that said 'outcome' is defined in terms of patients survival and/or relapse.

In this embodiment patients survival times and/or relapse are predicted according to their gene expression or genetic or epigenetic modifications thereof. By detecting patients with below average or below median metastasis free survival or disease free survival times and/or high likelihood of relapse the physician may choose to recommend the patient for further treatment, instead of or in addition to the endocrine targeting therapy(s), in particular but not limited to, chemotherapy.

The herein described invention provides a novel breast cell proliferative disorder prognostic and predictive biomarker.

It is herein described that aberrant expression of the gene PITX2 and/or regulatory or promoter regions thereof is correlated to prognosis and/or prediction of outcome of estrogen treatment of breast cell proliferative disorder patients, in particular breast carcinoma.

This marker thereby provides a novel means for the characterization of breast cell proliferative disorders. As described herein determination of the expression of the gene PITX2 and/or regulatory or promoter regions thereof enables the prediction of prognosis of a patient with a proliferative disorder of the breast tissues. In an alternative embodiment the expression of the gene PITX2 and/or regulatory or promoter regions thereof enables the prediction of treatment response of a patient treated with one or more treatments which target the estrogen receptor, synthesis or conversion pathways or are otherwise involved in estrogen metabolism, production or secretion.

The herein described invention is thereby useful for the differentiation of individuals who may be appropriately treated with one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion from those individuals, who would be optimally treated with other treatments in addition to said treatment. Preferred 'other treatments' include but are not limited to chemotherapy or radiotherapy. It is particularly preferred that said prognosis and/or treatment response is stated in terms of likelihood of relapse, survival or outcome.

In a further embodiment of the invention the aberrant expression of a plurality of genes comprising the gene PITX2 and/or regulatory or promoter regions thereof is analyzed. Said plurality of genes is hereinafter also referred to as a 'gene panel'. The analysis of multiple genes increases the accuracy of a provided prognosis and/or prediction of estrogen treatment outcome. It is preferred that the gene panel consists of up to seven genes and/or their promoter regions associated with prognosis and/or prediction of treatment response of breast carcinoma patients. It is further preferred that said panel consists of the gene PITX2 and one or more genes selected from the group consisting of ABCA8, CDK6, ERBB2, ONECUT2, PLAUI, TBC1D3 and TFF1 and/or regulatory regions thereof. It is particularly preferred that the gene panel is selected from the group of gene panels consisting of:

- PITX2, PLAUI and TFF1
- PITX2 and PLAUI
- PITX2 and TFF1

It is particularly preferred that the gene panel consisting PITX2 and TFF1 is used to predict outcome of treatment of patients with an endocrine treatment. It is particularly preferred that the gene panel consisting PITX2 and PLAUI is used to provide a prognosis of patients. It is preferred that said patients are analyzed prior to receiving any endocrine treatment.

In further embodiments this invention relates to new methods and sequences for the prognosis of patients diagnosed with breast cell proliferative disease. In a further aspect the invention relates to new methods and sequences, which may be used as tools for the selection of suitable treatments of patients diagnosed with breast cell proliferative disease based on a prediction of likelihood of relapse, survival or outcome.

More specifically this invention provides new methods and sequences for patients diagnosed with breast cell proliferative disease, allowing the improved selection of suitable adjuvant therapy. Furthermore, it is preferred that patients with poor prognosis following endocrine monotherapy are provided with chemotherapy in addition to or instead of an endocrine therapy.

One aspect of the invention is the provision of methods for providing a prognosis and/or prediction of outcome of endocrine treatment of a patient with a cell proliferative disorder of the breast tissues. Preferably said prognosis and/or prediction is provided in terms of likelihood of relapse or the survival of said patient. It is further preferred that said survival is disease free survival or metastasis free survival. It is also preferred that said disease is breast cancer. These methods comprise the analysis of the expression levels of the gene PITX2 and/or regulatory regions thereof.

In further embodiments the method comprises analysis of the expression of a 'gene panel' comprising the gene PITX2 and one or more genes selected from the group consisting of ABCA8, CDK6, ERBB2, ONECUT2, PLAUI, TBC1D3 and TFF1 and/or regulatory regions thereof. It is particularly preferred that said gene panels are selected from the group of gene panels consisting of:

- PITX2, PLAUI and TFF1
- PITX2 and PLAUI
- PITX2 and TFF1

It is particularly preferred that the expression of the gene panel consisting PITX2 and TFF1 is determined in order to predict outcome of treatment of patients with an endocrine treatment. It is also particularly preferred that the expression of the gene panel consisting PITX2 and PLAUI is determined in order to provide a prognosis of patients. It is preferred that said patients are analyzed prior to receiving any endocrine treatment.

Determination of expression may be achieved by any means standard in the art, however it is most preferably achieved by analysis of LOH, methylation, protein expression, mRNA expression, genetic or other epigenetic modifications of the genomic sequences.

Especially preferred is the analysis of the DNA methylation profile of the genomic sequence of the gene PITX2 and/or regulatory or promoter regions thereof as given in SEQ ID NO: 1130. Further preferred is the analysis of the methylation status of CpG positions within the following sections of SEQ ID NO: 1130 nucleotide 2,700-nucleotide 3,000; nucleotide 3,900-nucleotide 4,200; nucleotide 5,500-nucleotide 8,000; nucleotide 13,500-nucleotide 14,500; nucleotide 16,500-nucleotide 18,000; nucleotide 18,500-nucleotide 19,000; nucleotide 21,000-nucleotide 22,500. Especially preferred is the analysis of the methylation status of eight specific CpG dinucleotides, covered in the four sub-sequences of said SEQ ID NO: 1130 given in SEQ ID NOs: 23, 1140-1142. Wherein the method comprises analysis of a gene panel comprising the PITX2 and one or more genes selected from the group consisting ABCA8, CDK6, ERBB2, ONECUT2, PLAUI, TBC1D3 and TFF1 and/or regulatory or promoter regions thereof it is preferred that the sequence of said genes is selected from the group consisting of SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 AND SEQ ID NO: 1131 according to Table 1.

This methodology presents further improvements over the state of the art in that the method may be applied to any subject, independent of the estrogen and/or progesterone receptor status. Therefore in a preferred embodiment, the subject is not required to have been tested for estrogen or progesterone receptor status.

In further aspects of the invention, the disclosed matter provides novel nucleic acid sequences useful for the analysis of methylation within said gene, other aspects provide novel uses of the gene and the gene product as well as methods, assays and kits directed to providing a progn-

sis and/or predicting outcome of endocrine treatment of a patient diagnosed with breast cell proliferative disease.

In one embodiment the invention discloses a method for providing the prognosis and/or predicting outcome of endocrine treatment of a patient suffering from a breast cell proliferative disease, by analysis of expression of the gene PITX2 and/or regulatory regions thereof. Preferably said endocrine treatment is an adjuvant endocrine monotherapy. Said method may be enabled by means of any analysis of the expression of the gene, including but not limited to mRNA expression analysis or protein expression analysis or by analysis of its genetic modifications leading to an altered expression (including LOH). However, in the most preferred embodiment of the invention, said expression is determined by means of analysis of the methylation status of CpG sites within the gene PITX2 and its promoter or regulatory elements.

In one embodiment of the method aberrant expression of the gene PITX2 and/or panels thereof may be detected by analysis of loss of heterozygosity of the gene. In a first step genomic DNA is isolated from a biological sample of the patient's tumor. The isolated DNA is then analyzed for LOH by any means standard in the art including but not limited to amplification of the gene locus or associated microsatellite markers. Said amplification may be carried out by any means standard in the art including polymerase chain reaction (PCR), strand displacement amplification (SDA) and isothermal amplification.

The level of amplificate is then detected by any means known in the art including but not limited to gel electrophoresis and detection by probes (including Real Time PCR). Furthermore the amplicates may be labeled in order to aid said detection. Suitable detectable labels include but are not limited to fluorescence label, radioactive labels and mass labels the suitable use of which shall be described herein.

The detection of a decreased amount of an amplificate corresponding to one of the amplified alleles in a test sample as relative to that of a heterozygous control sample is indicative of LOH.

To detect the levels of mRNA encoding PITX2 and/or panels comprising said gene in a detection system for breast cancer relapse, a sample is obtained from a patient. Said obtaining of a sample is preferably not meant to be retrieving of a sample, as in performing a biopsy, but

rather directed to the availability of an isolated biological material representing a specific tissue, relevant for the intended use. The sample can be a tumor tissue sample from the surgically removed tumor, a biopsy sample as taken by a surgeon and provided to the analyst or a sample of blood, plasma, serum or the like. The sample may be treated to extract the nucleic acids contained therein. The resulting nucleic acid from the sample is subjected to gel electrophoresis or other separation techniques. Detection involves contacting the nucleic acids and in particular the mRNA of the sample with a DNA sequence serving as a probe to form hybrid duplexes. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd ed., 1989). Detection of the resulting duplex is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

In order to increase the sensitivity of the detection in a sample of mRNA encoding PITX2 and/or panels comprising said gene, the technique of reverse transcription/polymerization chain reaction can be used to amplify cDNA transcribed from mRNA encoding PITX2 and/or panels comprising said gene. The method of reverse transcription/PCR is well known in the art (for example, see Watson and Fleming, *supra*).

The reverse transcription/PCR method can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified using the PCR method and PITX2 and/or panels comprising said gene specific primers. (Belyavsky et al, Nucl Acid Res 17:2919-2932, 1989; Krug and Berger, Methods in Enzymology, Academic Press, N.Y., Vol.152, pp. 316-325, 1987 which are incorporated by reference)

The present invention may also be described in certain embodiments as a **kit** for use in predicting the likelihood of relapse and/or survival of a breast cancer patient before or after surgical tumor removal with or without adjuvant endocrine monotherapy state through testing of a biological sample. A representative kit may comprise one or more nucleic acid segments as described above that selectively hybridize to PITX2 mRNA and/or mRNA from genes of a panel comprising said PITX2 gene, and a container for each of the one or more nucleic acid segments. In certain embodiments the nucleic acid segments may be combined in a single tube. In further embodiments, the nucleic acid segments may also include a pair of primers for amplifying the target mRNA. Such kits may also include any buffers, solutions, solvents, enzymes, nucleotides, or other components for hybridization, amplification or detection reactions. Preferred kit components include reagents for reverse transcription-PCR, in situ hybridization, Northern analysis and/or RPA.

The present invention further provides for methods to detect the presence of the polypeptide(s) of, PITX2 and/or panels comprising said protein, in a sample obtained from a patient. It is preferred that said sequence is essentially the same as the sequence as given in Figure 107. Any method known in the art for detecting proteins can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. (for example see Basic and Clinical Immunology, Sites and Terr, eds., Appleton and Lange, Norwalk, Conn. pp 217-262, 1991 which is incorporated by reference). Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes of PITX2 and/or panels thereof and competitively displacing a labeled PITX2 protein and/or panels thereof or derivatives thereof.

Certain embodiments of the present invention comprise the use of antibodies specific to the polypeptide encoded by the gene PITX2 and/or panels comprising said gene. Such antibodies may be useful for providing a prognosis of the likelihood of relapse and/or survival of a breast cancer patient preferably under adjuvant endocrine monotherapy by comparing a patient's levels of PITX2 marker expression and/or the expression of panels comprising PITX2 to expression of the same marker(s) in normal individuals. In certain embodiments the production of monoclonal or polyclonal antibodies can be induced by the use of the PITX2 and/or other polypeptides of the panels as antigens. Such antibodies may in turn be used to detect expressed proteins as markers for prognosis of relapse of a breast cancer patient under adjuvant

endocrine monotherapy. The levels of such proteins present in the peripheral blood of a patient may be quantified by conventional methods. Antibody-protein binding may be detected and quantified by a variety of means known in the art, such as labeling with fluorescent or radioactive ligands. The invention further comprises kits for performing the above-mentioned procedures, wherein such kits contain antibodies specific for the PITX2 and/or panels thereof polypeptides.

Numerous competitive and non-competitive protein binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labeled for use a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like for use in radioimmunoassay (RIA), enzyme immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like. Polyclonal or monoclonal antibodies to PITX2 and/or panels thereof or an epitope thereof can be made for use in immunoassays by any of a number of methods known in the art. One approach for preparing antibodies to a protein is the selection and preparation of an amino acid sequence of all or part of the protein, chemically synthesising the sequence and injecting it into an appropriate animal, usually a rabbit or a mouse (Milstein and Kohler Nature 256:495-497, 1975; Gutfre and Milstein, Methods in Enzymology: Immunochemical Techniques 73:1-46, Langone and Banatis eds., Academic Press, 1981 which are incorporated by reference). Methods for preparation of PITX2 and/or panels thereof or an epitope thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples.

In one aspect the invention provides significant improvements over the state of the art in that it is the first single marker that can be used to predict the likelihood of relapse or of survival of a breast cancer patient under adjuvant endocrine monotherapy.

In the most preferred embodiment of the invention the analysis of expression is carried out by means of methylation analysis. It is further preferred that the methylation state of the CpG dinucleotides within the genomic sequence according to SEQ ID NO: 1130 and sequences complementary thereto is analyzed. SEQ ID NO: 1130 discloses the gene PITX2 and its promoter and regulatory elements thereof, wherein said fragment comprises CpG dinucleotides exhibiting a prognosis and/or predicting outcome of endocrine treatment specific methylation

pattern. Further preferred is the analysis of the methylation status of CpG positions within the following sections of SEQ ID NO: 1130 nucleotide 2,700-nucleotide 3,000; nucleotide 3,900-nucleotide 4,200; nucleotide 5,500-nucleotide 8,000; nucleotide 13,500-nucleotide 14,500; nucleotide 16,500-nucleotide 18,000; nucleotide 18,500-nucleotide 19,000; nucleotide 21,000-nucleotide 22,500. Also preferred is the analysis of the sub-sequence of the gene PITX2 as shown in SEQ ID NO: 23.

Wherein the method comprises analysis of the expression of a 'gene panel' comprising the gene and/or regulatory or promoter regions thereof and one or more genes selected from the group consisting ABCA8, CDK6, ERBB2, ONECUT2, PLAU, TBC1D3 and TFF1 it is almost most preferred that said analysis of expression is carried out by means of methylation analysis. It is particularly preferred that the CpG methylation of the gene panels selected from the group of gene panels consisting:

- PITX2, PLAU and TFF1
- PITX2 and PLAU
- PITX2 and TFF1

is analyzed.

It is particularly preferred that the methylation of the gene panel consisting PITX2 and TFF1 is determined in order to predict outcome of treatment of patients with an endocrine treatment. It is also particularly preferred that the methylation of the gene panel consisting PITX2 and PLAU is determined in order to provide a prognosis of patients. It is preferred that said patients are analyzed prior to receiving any endocrine treatment.

Hypermethylation of PITX2 and selected other genes as herein and/or sequences thereof are associated with poor prognosis and/or outcome of endocrine treatment of breast cell proliferative disorders, most preferably breast carcinoma.

The methylation pattern of the gene PITX2 and its promoter and regulatory elements have heretofore not been analyzed with regard to prognosis or prediction of outcome of endocrine treatment of a patient diagnosed with a breast cell proliferative disorder. Due to the degeneracy of the genetic code, the sequence as identified in SEQ ID NO: 1130 should be interpreted so as to include all substantially similar and equivalent sequences upstream of the promoter

region of a gene which encodes a polypeptide with the biological activity of that encoded by PITX2.

Most preferably, the following method is used to detect methylation within the gene PITX2 and/or regulatory or promoter regions thereof wherein said methylated nucleic acids are present in an excess of background DNA, wherein the background DNA is present in 100 to 1000 times the concentration of the DNA to be detected.

The method for the analysis of methylation comprises contacting a nucleic acid sample obtained from a subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

Preferably, said method comprises the following steps: In the first *step*, a sample of the tissue to be analyzed is obtained. The source may be any suitable source, preferably, the source of the sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, plasma, serum, stool, urine, blood, nipple aspirate and combinations thereof. Preferably, the source is tumor tissue, biopsies, serum, urine, blood or nipple aspirate. The most preferred source, is the tumor sample, surgically removed from the patient or a biopsy sample of said patient.

The DNA is then isolated from the sample. Genomic DNA may be isolated by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in/by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

The genomic DNA sample is then treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dis-

similar to cytosine in terms of hybridization behavior. This will be understood as "treatment" or "pre-treatment" herein.

The above described pre-treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior. Enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996) is one preferred example how to perform said pre-treatment. It is further preferred that the bisulfite treatment is carried out in the presence of a radical scavenger or DNA denaturing agent.

The treated DNA is then analyzed in order to determine the methylation state of the gene PITX2 and/or regulatory regions thereof (prior to the treatment) associated with prognosis and/or outcome of endocrine treatment. In a further embodiment of the method the methylation state of the gene PITX2 and/or regulatory regions thereof and the methylation state of one or more genes selected from the group consisting ABCA8, CDK6, ERBB2, ONECUT2, PLAU, TBC1D3 and TFF1 and/or regulatory or promoter regions thereof is determined. It is particularly preferred that methylation status of a gene panel selected from the group of gene panels consisting PITX2, PLAU and TFF1; PITX2 and PLAU; PITX2 and TFF1 is determined. It is further preferred that the sequences of said genes as described in the accompanying sequence listing (see Table 3) are analyzed.

In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said amplificates comprise at least one 20 base pair sequence comprising at least three CpG dinucleotides. Said amplification is carried out using sets of primer oligonucleotides according to the present invention, and a preferably heat-stable polymerase. The amplification of several DNA segments can be carried out simultaneously in one and the

same reaction vessel, in one embodiment of the method preferably six or more fragments are amplified simultaneously. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 18-base-pair long segment of the base sequences of SEQ ID NO: 250-251, 372-373, SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 AND SEQ ID NO: 1132 to SEQ ID NO: 1139 and sequences complementary thereto.

In a preferred embodiment of the method the primers may be selected from the group consisting to SEQ ID NO: 1143 to SEQ ID NO: 1147.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising SEQ ID NO: 23, SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12, SEQ ID NO: 1130 and SEQ ID NO: 1131 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG, TpG or CpA dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 18 nucleotides which hybridizes to a pretreated nucleic acid sequence according to SEQ ID NO: 250-251, 372-373 and SEQ ID NO: 1132, 1133, 1136 and 1137 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, tpG or Cpa dinucleotide. In this embodiment of the method according to the invention it is particularly preferred that the MSP primers comprise between 2 and 4 CpG, tpG or Cpa dinucleotides. It is further preferred that said dinucleotides are located within the 3' half of the primer e.g. wherein a primer is 18 bases in length the specified dinucleotides are located within the first 9 bases from the 3' end of the molecule. In addition to the CpG, tpG or Cpa dinucleotides it is further preferred that said primers should further comprise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridizing strand, guanine converted to adenosine). In a further pre-

ferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, *e.g.*, matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas and Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapor phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut and Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut and Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of

charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In a particularly preferred embodiment of the method the amplification of step three is carried out in the presence of at least one species of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. The use of blocking oligonucleotides enables the improved specificity of the amplification of a sub-population of nucleic acids. Blocking probes hybridized to a nucleic acid suppress, or hinder the polymerase mediated amplification of said nucleic acid. In one embodiment of the method blocking oligonucleotides are designed so as to hybridize to background DNA. In a further embodiment of the method said oligonucleotides are designed so as to hinder or suppress the amplification of unmethylated nucleic acids as opposed to methylated nucleic acids or vice versa.

Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'TpG' at the position in question, as opposed to a 'CpG.' In one embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to molecule is complementary or identical to a sequence at least 18 base pairs in length selected from the group consisting of SEQ ID NOs: 250-251, 372-373, 1132, 1133, 1136 and 1137 preferably comprising one or more CpG, TpG or CpA dinucleotides. In one embodiment of the method the sequence of said oligonucleotides is selected from the group consisting SEQ ID NO: 1148 and SEQ ID NO: 1149 and sequences complementary thereto.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivatised at the 3' position with other than a "free" hydroxyl group. For ex-

ample, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (e.g., with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker - a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

In one embodiment of the method, the binding site of the blocking oligonucleotide is identical to, or overlaps with that of the primer and thereby hinders the hybridization of the primer to its binding site. In a further preferred embodiment of the method, two or more such blocking oligonucleotides are used. In a particularly preferred embodiment, the hybridization of one of the blocking oligonucleotides hinders the hybridization of a forward primer, and the hybridization of another of the probe (blocker) oligonucleotides hinders the hybridization of a reverse primer that binds to the amplificate product of said forward primer.

In an alternative embodiment of the method, the blocking oligonucleotide hybridizes to a location between the reverse and forward primer positions of the treated background DNA, thereby hindering the elongation of the primer oligonucleotides.

It is particularly preferred that the blocking oligonucleotides are present in at least 5 times the concentration of the primers.

In the fourth step of the method, the amplicates obtained during the third step of the method are analyzed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplicates are obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplicate is in itself indicative of the methylation state of the CpG positions covered by the primers and or blocking oligonucleotide, according to the base sequences thereof. All possible known molecular biological methods may be used for this detection, including, but not limited to gel electrophoresis, sequencing, liquid chromatography, hybridizations, real time PCR analysis or combinations thereof. This step of the method further acts as a qualitative control of the preceding steps.

In the fourth step of the method amplicates obtained by means of both standard and methylation specific PCR are further analyzed in order to determine the CpG methylation status of the genomic DNA isolated in the first step of the method. This may be carried out by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplicates synthesized in step three are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplicates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the SEQ ID NO: 250-251, 372-373 and SEQ ID Nos: 1132, 1133, 1136 and 1137 and the segment comprises at least one CpG, TpG or CpA dinucleotide. In further embodiments said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-

397, 358-359, 412-413, 350-351; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. In a further embodiment one oligonucleotide exists for the analysis of each CpG dinucleotide within the sequences according to SEQ ID NO: 23 and 1130, and the equivalent positions within SEQ ID NO: 250-251, 372-373 and SEQ ID NO: 1132, 1133, 1136 and 1137. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 23, SEQ ID NOS. 1130, 1131, AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12, and the equivalent positions within SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139, AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplicates are then removed. The hybridized amplicates are detected. In this context, it is preferred that labels attached to the amplicates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridized to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393). There are two preferred embodiments of utilizing this method. One embodiment, known as the TaqMan™ assay employs a dual-labeled fluorescent oligonucleotide probe. The TaqMan™ PCR reaction employs the use of a non-extendible interrogating oligonucleotide, called a TaqMan™ probe, which is designed to hybridize to a CpG-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oli-

gonucleotide. Hybridized probes are displaced and broken down by the polymerase of the amplification reaction thereby leading to an increase in fluorescence. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight assay. The second preferred embodiment of this MethyLight technology is the use of dual-probe technology (Lightcycler®), each probe carrying donor or recipient fluorescent moieties, hybridization of two probes in proximity to each other is indicated by an increase or fluorescent amplification primers. Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

Also any combination of these probes or combinations of these probes with other known probes may be used.

In a further preferred embodiment of the method, the fourth step of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997. In said embodiment it is preferred that the methylation specific single nucleotide extension primer (MS-SNuPE primer) is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the sequences taken from the group of SEQ ID NO: 250-251, 372-373 and SEQ ID Nos: 1132, 1133, 1136 and 1137. However it is preferred to use fluorescently labeled nucleotides, instead of radiolabeled nucleotides.

In yet a further embodiment of the method, the fourth step of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the third step of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

In the most preferred embodiment of the methylation analysis method the genomic nucleic acids are isolated and treated according to the first three steps of the method outlined above, namely:

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
- b) extracting or otherwise isolating the genomic DNA;

- c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties; and wherein
- d) amplifying subsequent to treatment in c) is carried out in a methylation specific manner, namely by use of methylation specific primers or blocking oligonucleotides, and further wherein
- e) detecting of the amplicates is carried out by means of a real-time detection probe, as described above.

Preferably, where the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 250-251, 372-373 and SEQ ID Nos: 1132, 1133, 1136 and 1137 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide. Additionally, further methylation specific primers may also be used for the analysis of a gene panel as described above wherein said primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 and SEQ ID Nos: 1134, 1135, 1138 and 1139 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

In an alternative most preferred embodiment of the method, the subsequent amplification of d) is carried out in the presence of blocking oligonucleotides, as described above. It is particularly preferred that said blocking oligonucleotides comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 250-251, 372-373, SEQ ID Nos: 1132, 1133, 1136 and 1137 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide.

Additionally, further blocking oligonucleotides may also be used for the analysis of a gene panel as described above wherein said blocking oligonucleotides comprising a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-

229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 and SEQ ID Nos: 1134, 1135, 1138 and 1139 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide.

Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139, AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351, and most preferably SEQ ID NO: 250-251, 372-373 and SEQ ID Nos: 1132, 1133, 1136 and 1137 is carried out by means of real-time detection methods as described above.

Additional embodiments of the invention provide a method for the analysis of the methylation status of the gene PITX2 and/or regulatory regions thereof without the need for pre-treatment. Furthermore said method may also be used for the methylation analysis of the gene PITX2 and/or regulatory regions thereof and the methylation state of one or more genes selected from the group consisting ABCA8, CDK6, ERBB2, ONECUT2, PLAUI, TBC1D3, TFF1 and/or regulatory or promoter regions thereof is determined. It is particularly preferred that methylation status of a gene panel selected from the group of gene panels consisting PITX2, PLAUI and TFF1; PITX2 and PLAUI; PITX2 and TFF1 is determined.

In the first step of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, biopsy tissue, body fluids, or breast tumor tissue embedded in paraffin. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, but preferably with methylation-sensitive restriction endonucleases.

In the second step, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the third step, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplicates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

In the fourth step the amplicates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

In the final step of the method the prognosis and/or predicting outcome of endocrine treatment is determined. Preferably, the correlation of the expression level of the genes with the prognosis and/or predicting outcome of endocrine treatment is done substantially without human intervention. Poor prognosis and/or predicting outcome of endocrine treatment is determined by aberrant levels of mRNA and/or protein, and hypermethylation. It is particularly preferred that said hypermethylation is above average or above median of said disease in said specific setting.

It is particularly preferred that the classification of the sample is carried out by algorithmic means.

In one embodiment machine learning predictors are trained on the methylation patterns at the investigated CpG sites of the samples with known status. A selection of the CpG positions which are discriminative for the machine learning predictor are used in the panel. In a particularly preferred embodiment of the method, both methods are combined; that is, the machine learning classifier is trained only on the selected CpG positions that are significantly differentially methylated between the classes according to the statistical analysis.

The development of algorithmic methods for the classification of a sample based on the methylation status of the CpG positions within the panel are demonstrated in the examples.

The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof. In a preferred embodiment of the invention, the objective comprises analysis of a non-naturally occurring modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12, wherein said sequence comprises at least one CpG, TpA or CpA dinucleotide and sequences complementary thereto. The sequences of SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 provide non-naturally occurring modified versions of the nucleic acid according to SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12, wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic sequence as follows. For each sense strand genomic DNA, e.g., SEQ ID NO: 23, four converted versions are disclosed. A first version wherein "C" to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 correspond to SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-

359, 412-413, 350-351 . A third chemically converted version of each genomic sequences is provided, wherein "C" to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are unmethylated); a final chemically converted version of each sequence, discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the complement (antisense strand) of each genomic sequence, all "C" residues of CpG dinucleotide sequences are unmethylated). The 'downmethylated' converted sequences of SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 correspond to SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 .

The invention further discloses oligonucleotide or oligomer for detecting the cytosine methylation state within genomic or pre-treated DNA, according to SEQ ID NO: 23, SEQ ID NO: 1130 to SEQ ID NO: 1139 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a treated nucleic acid sequence according to SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337,

396-397, 358-359, 412-413, 350-351, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (e.g., a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or T_m , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the T_m , the temperature of the final wash in the hybridisation reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in T_m can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO:23, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X, where

the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to $(n + (X-1))$;

where $n=1, 2, 3, \dots (Y-(X-1))$;

where Y equals the length (nucleotides or base pairs) of SEQ ID NO: 23 (9001);

where X equals the common length (in nucleotides) of each oligonucleotide in the set (e.g., $X=20$ for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to $Y-(X-1)$. For example $Z=9001-19=8,982$ for either sense or antisense sets of SEQ ID NO: 23, where $X=20$.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Examples of inventive 20-mer oligonucleotides include the following set of oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO: 23: 1-20, 2-21, 3-22, 4-23, 5-24, and 8,982 –9,001.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Likewise, examples of inventive 25-mer oligonucleotides include the following set of oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO: 23: 1-25, 2-26, 3-27, 4-28, 5-29,and 8,977–9,001.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

The present invention encompasses, for each of SEQ ID NO: 23, -250-251, , 372-373, SEQ ID NO: 1130 to AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, e.g., $X=9, 10, 17, 20, 22, 23, 25, 27, 30$ or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO: 23, 250-251, 372-373, SEQ ID NO: 1130 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophores, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *Bio-Techniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a chromophore, fluorophore, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequences SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the treated nucleic acids according to SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 and sequences complementary thereto. However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state of treated genomic DNA (SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351), or in genomic DNA (SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 and sequences complementary thereto). These probes enable diagnosis, and/or classification of genetic and epigenetic parameters of lung cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in treated genomic DNA (SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351), or in genomic DNA (SEQ ID NO: 23, SEQ ID NO: 1130, SEQ ID NO: 1131 AND SEQ ID NO: 49, SEQ ID NO: 46, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 16, SEQ ID NO: 43, SEQ ID NO: 12 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NO: 250-251, 372-373, SEQ ID NO: 1132 to SEQ ID NO: 1139 AND SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (i.e., an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form of pellets or also as resin matrices, may also be used. An overview of the prior art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999, and from the literature cited therein). Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

It is also anticipated that the oligonucleotides, or particular sequences thereof, may constitute all or part of a "virtual array" wherein the oligonucleotides, or particular sequences thereof, are used, for example, as 'specifiers' as part of, or in combination with a diverse population of unique labeled probes to analyze a complex mixture of analytes. Such a method, for example is described in US 2003/0013091 (United States serial number 09/898,743, published 16 January 2003). In such methods, enough labels are generated so that each nucleic acid in the complex mixture (i.e., each analyte) can be uniquely bound by a unique label and thus detected (each label is directly counted, resulting in a digital read-out of each molecular species in the mixture).

The described invention further provides a composition of matter useful for providing a prognosis and/or prediction of outcome of endocrine treatment of breast cancer patients. Said

composition comprising at least one nucleic acid 18 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 250-251, 372-373, 1132, 1133, 1136 and 1137, and one or more substances taken from the group comprising : magnesium chloride, dNTP, taq polymerase, bovine serum albumen, an oligomer in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridizes under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 250-251, 372-373 and SEQ ID NO: 1132, 1133, 1136 and 1137 and sequences complementary thereto. It is preferred that said composition of matter comprises a buffer solution appropriate for the stabilization of said nucleic acid in an aqueous solution and enabling polymerase based reactions within said solution. Suitable buffers are known in the art and commercially available.

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent as well as at least one oligonucleotide whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 18-base long segment of the sequences SEQ ID NO: 250-251, 372-373, 1132, 1133, 1136 and 1137. Said kit may further comprise at least one oligonucleotide whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 18-base long segment of the sequences SEQ ID Nos: 302-303, 296-297, 214-215, 274-275, 236-237, 290-291, 228-229, 250-251, 424-425, 418-419, 336-337, 396-397, 358-359, 412-413, 350-351. Said kit may further comprise instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight®, HeavyMethyl®, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

Typical reagents (*e.g.*, as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligonucleotide probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation

buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MethyLight®-based kit) for MethyLight® analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples and figures serve only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples and figures serve only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

Figure 1 shows a preferred application of the method according to the invention. The X axis shows the tumour(s) mass, wherein the line '3' shows the limit of detectability. The Y-axis

shows time. Accordingly said figure illustrates a simplified model of endocrine treatment of an Stage 1-3 breast tumour wherein primary treatment was surgery (at point 1), followed by adjuvant therapy with Tamoxifen. In a first scenario a responder to treatment (4) is shown as remaining below the limit of detectability for the duration of the observation. A non responder to the treatment (5) has a period of disease free survival (2) followed by relapse when the carcinoma mass reaches the level of detectability.

Figure 2 shows another preferred application of the method according to the invention. The X axis shows the tumour(s) mass, wherein the line '3' shows the limit of detectability. The Y-axis shows time. Accordingly said figure illustrates a simplified model of Endocrine treatment of an late stage breast tumour wherein primary treatment was surgery (at point 1), followed by relapse which is treated by Tamoxifen (2). In a first scenario a responder to treatment (4) is shown as remaining below the limit of detectability for the duration of the observation. A non responder to the treatment (5) does not recover from the relapse.

Figures 3 to 45 show the Kaplan-Meier estimated disease-free survival curves for single genes or oligonucleotide positions. The black plot shows the proportion of disease free patients in the population with above median methylation levels, the grey plot shows the proportion of disease free patients in the population with below median methylation levels

Figure 46 shows the methylation analysis of CpG islands according to Example 1. CpG islands per gene were grouped and their correlation with objective response determined by Hotelling's T^2 statistics. Black dots indicate the P -value of the indicated gene. The 20 most informative genes, ranked from left to right with increasing P -value, are shown. The top dotted line marks the uncorrected significance value ($P < 0.05$). The lower dotted line marks significance after false discovery rate correction of 25%. All genes with a P -value smaller or equal to the gene with the largest P -value that is below the lower line (in this case COX7A2L) are considered significant. The FDR correction chosen guarantees that the identified genes are with 75% chance true discoveries.

Figures 48 shows a ranked matrix of the best 11 amplicates of data obtained according to Example 1 (Metastatic setting, limited sample set). P -values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models. The figure is shown in grey-scale, wherein the most significant CpG positions are at the bottom of the matrix with signifi-

cance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in grey, from light (low proportion of methylation) to dark (high proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the methylation profile for the different CpGs for one sample. The p-values for the individual CpG positions are shown on the right side. The p-values are the probabilities that the observed distribution occurred by chance in the data set.

Figure 49 shows a ranked matrix of some of the best markers obtained according to Example 1 (Metastatic setting, limited sample set). P-values were calculated from Likelihood ratio (LR) tests from univariate logistic regression models. The figure is shown in greyscale, wherein the most significant CpG positions are at the bottom of the matrix with significance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in grey, from light (low proportion of methylation) to dark (high proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the methylation profile for the different CpGs for one sample. The p-values for the individual CpG positions are shown on the right side. The p-values are the probabilities that the observed distribution occurred by chance in the data set.

Figures 47 and 50 show the uncorrected p-values on a log-scale. P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models according to Example 1 (metastatic setting). Each individual genomic region of interest is represented as a point, the upper dotted line represents the cut off point for the 25% false discovery rate, the lower dotted line shows the Bonferroni corrected 5% limit.

Figure 51 shows a ranked matrix of the best 11 amplificates of data obtained according to Example 1 (Metastatic setting, all samplews). P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models. The figure is shown in greyscale, wherein the most significant CpG positions are at the bottom of the matrix with significance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in grey, from light (low proportion of methylation) to dark (high proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the

methylation profile for the different CpGs for one sample. The p-values for the individual CpG positions are shown on the right side. The p-values are the probabilities that the observed distribution occurred by chance in the data set.

Figure 52 shows the disease-free survival curves for a combination of two oligonucleotides each from the genes TBC1D3 and CDK6, and one oligonucleotide from the gene PITX2. The black plot shows the proportion of disease free patients in the population with above median methylation levels, the grey plot shows the proportion of disease free patients in the population with below median methylation levels.

Figure 53 shows the plot according to Figure 52 and the classification of the sample set by means of the St. Gallen method. The unbroken lines represent the methylation analysis wherein the black plot shows the proportion of disease free patients in the population with above median methylation levels, the grey plot shows the proportion of disease free patients in the population with below median methylation levels. The broken lines represent the St. Gallen classification of the sample set wherein the black plot shows the disease free survival time of the *high risk group* and the grey plot shows the disease free survival of the *low risk group*.

Figure 54 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis. The lower plot shows the proportion of disease free patients in the population with above median methylation levels, the upper plot shows the proportion of disease free patients in the population with below median methylation levels. The X axis shows the disease free survival times of the patients in months, and the Y- axis shows the proportion of disease free survival patients.

Figure 55 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the ERBB2 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 56 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the ERBB2 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 57 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the TFF1 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 58 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the TFF1 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of metastasis free patients in the population with below an optimised cut off point's methylation levels.

Figure 59 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PLAU gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 60 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the PLAU gene by means of Real-Time methylation specific probe analysis according

to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of metastasis free patients in the population with below an optimised cut off point's methylation levels.

Figure 61 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 62 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 63 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the TBC1D3 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 64 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the TBC1D3 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The

black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 65 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the **ERBB2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 66 shows the Kaplan-Meier estimated **metastasis-free** survival curves for a CpG position of the **ERBB2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 67 shows the Kaplan-Meier estimated **disease-free** survival curves for a CpG position of the **TFF1** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 68 shows the Kaplan-Meier estimated **metastasis-free** survival curves for a CpG position of the **TFF1** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off

point's methylation levels, the grey plot shows the proportion of metastasis free patients in the population with below an optimised cut off point's methylation levels.

Figure 69 shows the Kaplan-Meier estimated **disease-free survival** curves for a CpG position of the **PLAU** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 70 shows the Kaplan-Meier estimated **metastasis-free survival** curves for a CpG position of the **PLAU** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 71 shows the Kaplan-Meier estimated **disease-free survival** curves for a CpG position of the **PITX2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 72 shows the Kaplan-Meier estimated **metastasis-free survival** curves for a CpG position of the **PITX2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of metastasis free patients in the population with below an optimised cut off point's methylation levels.

Figure 73 shows the Kaplan-Meier estimated **disease-free** survival curves for a CpG position of the **PITX2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 74 shows the Kaplan-Meier estimated **metastasis-free** survival curves for a CpG position of the **PITX2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 75 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the **ONECUT2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 76 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the **ONECUT2** gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 77 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the TBC1D3 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 78 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the TBC1D3 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 79 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the ABCA8 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 80 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of the ABCA8 gene by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 81 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the TFF1 (SEQ ID NO: 12) and PLAUI (SEQ ID NO:16) genes by means

of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 82 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the **TFF1** (SEQ ID NO: 12) and **PLAU** (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 83 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the **TFF1** (SEQ ID NO: 12) and **PLAU** (SEQ ID NO:16) and **PITX2** (SEQ ID NO:23) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 84 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the **TFF1** (SEQ ID NO: 12) and **PLAU** (SEQ ID NO:16) and **PITX2** (SEQ ID NO:23) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 85 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the PITX2 (SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 86 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position a combination of the PITX2 (SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 87 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the PITX2 (SEQ ID NO:23) and PLAU (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 88 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the PITX2 (SEQ ID NO:23) and PLAU (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation

levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 89 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 90 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 91 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2 (SEQ ID NO:23) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 92 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2 (SEQ ID NO:23) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and

the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 93 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the **PITX2** (SEQ ID NO:23) and **TFF1** (SEQ ID NO: 12) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 94 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the **PITX2** (SEQ ID NO:23) and **TFF1** (SEQ ID NO: 12) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the metastasis free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 95 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of a combination of the **PITX2** (SEQ ID NO:23) and **PLAU** (SEQ ID NO:16) genes by means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 96 shows the Kaplan-Meier estimated metastasis-free survival curves for a CpG position of a combination of the **PITX2** (SEQ ID NO:23) and **PLAU** (SEQ ID NO:16) genes by

means of Real-Time methylation specific probe analysis according to Example 2. The X axis shows the disease free survival times of the patients in years, and the Y- axis shows the proportion of patients with metastasis free survival. The black plot shows the proportion of metastasis free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

Figure 97 shows a scatter plot of matched pair PET and fresh frozen tissues analysed using PITX2 gene assay 1 according to Example 2. Quantitative methylation CT scores of PET samples are shown on the Y-axis, and quantitative methylation CT scores of fresh frozen samples are shown on the X-axis. The association between the paired samples is 0.81 (Spearman's rho). This analysis is based on n=89 samples.

Figure 98 shows the Disease free survival (DFS) of randomly selected ER+, N0, untreated patient population in Kaplan-Meier survival plot according to Example 2. Proportion of disease free patients is shown on the Y-axis and time in years is shown on the X-axis. 139 events were observed (observed event rate=33%). Disease free survival after 5 years: 74.5% [70.3%, 78.9%], after 10 years 59.8% [54.2%, 66%]. 95% confidence intervals are plotted.

Figure 99 shows the distribution of follow-up times in ER+, N0, untreated population according to Example 2. Frequency is shown on the Y-axis and time in months is shown on the X-axis. The figure on the left shows patients with event (all kinds of relapses). Mean follow-up time 45.8 months (standard deviation=31), median=38 (range=[2, 123]).

The figure on the right shows censored patients. Mean follow up time 93 months (standard deviation=35.6), median=94 (range=[1, 190]).

Figure 100 shows the Disease free survival (DFS) of ER+, N0, TAM treated population in Kaplan-Meier plot according to Example 2. Proportion of disease free patients is shown on the Y-axis and time in years is shown on the X-axis. 56 events were observed (observed event rate=10 %). DFS after 5 years: 92.4% [90%, 94.9%], after 10 years: 82.1% [77.3%,87.2%]. 95% confidence intervals are plotted.

Figure 101 shows the distribution of follow-up times in ER+, N0, untreated population according to Example 2. Frequency is shown on the Y-axis and time in months is shown on the

X-axis. The figure on the left shows patients with all events (all kinds of relapses). Mean follow-up time 47.9 months (standard deviation=24.4), median=45 (range=[2, 98]).

The figure on the right shows censored patients. Mean follow up time 65.3 months (standard deviation=31.6), median=64 (range=[0, 158]).

Figure 102 shows the ROC plot at different times for marker model 3522 (Assay 1) and 2265 on ER+N0 TAM treated population according to Example 2. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates, and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.32 quantile) are plotted.

Figure 103 shows the ROC plot at different times for marker model 3522 (Assay 1) alone on ER+N0 TAM treated population according to Example 2. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates, and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.42 quantile) are plotted.

Figure 104 shows the ROC plot at different times for marker model 2265 on ER+N0 TAM treated population according to Example 2. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates for different thresholds (= 5, 6, 7, 8 years) and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.78 quantile) are plotted.

Figure 105 shows the ROC plot at different times for marker model 2395 on ER+N0 TAM treated population according to Example 2. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates for different thresholds (= 5, 6, 7, 8 years), and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.77 quantile) are plotted.

SEQ ID NOS: 1 to 61 and 149 to 150 represent 5' and/or regulatory regions and/or CpG rich regions of the genes according to Table 1. These sequences are derived from Genbank and will be taken to include all minor variations of the sequence material which are currently unforeseen, for example, but not limited to, minor deletions and SNPs.

Example 1

DNA samples were extracted using the Wizzard Kit (Promega), samples from 278 patients were analysed, data analyses were carried out on a selection of candidate markers.

Bisulfite treatment and mPCR

Total genomic DNA of all samples was bisulfite treated converting unmethylated cytosines to uracil. Methylated cytosines remained conserved. Bisulfite treatment was performed with minor modifications according to the protocol described in Olek et al. (1996). After bisulfitation 10 ng of each DNA sample was used in subsequent mPCR reactions containing 6-8 primer pairs.

Each reaction contained the following:

2.5 pmol each primer

11.25 ng DNA (bisulfite treated)

Multiplex PCR Master mix (Qiagen)

Further details of the primers are shown in TABLE 2.

Initial denaturation was carried out at 95°C for 15 min. Forty cycles were carried out as follows: Denaturation at 95°C for 30 sec, followed by annealing at 57°C for 90 sec., primer elongation at 72°C for 90 sec. A final elongation at 72°C was carried out for 10 min.

Hybridisation

All PCR products from each individual sample were then hybridised to glass slides carrying a pair of immobilised oligonucleotides for each CpG position under analysis. Each of these detection oligonucleotides was designed to hybridise to the bisulphite converted sequence around one CpG site which was either originally unmethylated (TG) or methylated (CG). See Table 2 for further details of hybridisation oligonucleotides used. Hybridisation conditions were selected to allow the detection of the single nucleotide differences between the TG and CG variants.

5 µl volume of each multiplex PCR product was diluted in 10 x Ssarc buffer . The reaction mixture was then hybridised to the detection oligonucleotides as follows. Denaturation at 95°C, cooling down to 10 °C, hybridisation at 42°C overnight followed by washing with 10 x Ssarc and dH₂O at 42°C. Further details of the hybridisation oligonucleotides are shown in TABLE 3.

Fluorescent signals from each hybridised oligonucleotide were detected using genepix scanner and software. Ratios for the two signals (from the CG oligonucleotide and the TG oligonucleotide used to analyse each CpG position) were calculated based on comparison of intensity of the fluorescent signals.

Data analysis methods

Analysis of the chip data: From raw hybridisation intensities to methylation ratios; The log methylation ratio ($\log(\text{CG}/\text{TG})$) at each CpG position is determined according to a standardised preprocessing pipeline that includes the following steps: For each spot the median background pixel intensity is subtracted from the median foreground pixel intensity (this gives a good estimate of background corrected hybridisation intensities); For both CG and TG detection oligonucleotides of each CpG position the background corrected median of the 4 redundant spot intensities is taken; For each chip and each CpG position the $\log(\text{CG}/\text{TG})$ ratio is calculated; For each sample the median of $\log(\text{CG}/\text{TG})$ intensities over the redundant chip

repetitions is taken. This ratio has the property that the hybridisation noise has approximately constant variance over the full range of possible methylation rates (Huber et al., 2002).

Hypothesis testing

The main task is to identify markers that show significant differences in the average degree of methylation between two classes. A significant difference is detected when the null hypothesis that the average methylation of the two classes is identical can be rejected with $p < 0.05$. Because we apply this test to a whole set of potential markers we have to correct the p-values for multiple testing. This was done by applying the False Discovery Rate (FDR) method (Dudoit et al., 2002).

For testing the null hypothesis that the methylation levels in the two classes are identical we used the likelihood ratio test for logistic regression models (Venables and Ripley, 2002). The logistic regression model for a single marker is a linear combination of methylation measurements from all CpG positions in the respective genomic region of interest (ROI). A significant p-value for a marker means that this ROI has some systematic correlation to the question of interest as given by the two classes. However, at least formally it makes no statement about the actual predictive power of the marker.

Logistic Regression

Logistic regression models are tools to model the probability of an event in dependence of one or more variables or factors. For example, if x denotes a specific methylation log ratio, the probability that a patient responds to the applied therapy (Tamoxifen) is modeled as

$$P(\text{response} \mid x) = \exp(\alpha + \beta x) / [1 + \exp(\alpha + \beta x)]. \quad (1)$$

If x_1, \dots, x_k denote the k methylation logratios measured for one amplificate, the model is

$$P(\text{response} \mid x_1, \dots, x_k) = \exp(\alpha + \beta_1 x_1 + \dots + \beta_k x_k) / [1 + \exp(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)]. \quad (2)$$

Significance of the respective amplificate is assessed using a likelihood-ratio test. This test calculates the difference of $-2\text{Log}(\text{likelihood})$ for the full model and the null-model including just the intercept α which is approximately χ^2 -distributed with k degrees of freedom under the null hypotheses $\beta_1 = \dots = \beta_k = 0$.

If additional covariates are considered, the model contains an additional parameter for each covariate and the test statistic is calculated as the difference of $-2\text{Log}(\text{likelihood})$ or the full model and the null-model including intercept and covariates. Again, given the null hypothesis, this difference is approximately χ^2 -distributed with k degrees of freedom.

Ranked Matrices

For a graphical display of all group comparisons, ranked matrices are used. Each row represents one oligo pair, whereas each column of the matrix stands for one sample (or chip in the case of up- versus downmethylated Promega DNA comparisons). Oligo pairs are ranked according to their discriminatory power (Wilcoxon test, Fisher score or logistic regression), where the best “marker” is displayed on the bottom line. Low methylation is displayed in light grey, high methylation in dark grey, and the data are normalized prior to display.

Cox Regression

Disease-free survival times (DFS) are modeled using Cox regression models. These models are similar to logistic regression models, but instead of probabilities, the hazard is modeled. The hazard gives the instantaneous risk of a relapse. The models

$$h(t | x) = h_0(t) \cdot \exp(\beta x) \quad (3)$$

and

$$h(t | x_1, \dots, x_k) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_k x_k) \quad (4)$$

are used for uni- and multivariate analyses, respectively, where t is the time measured in months after surgery and $h_0(t)$ is the baseline hazard. Likelihood ratio tests are performed similar to those used for logistic regression. Again, the difference between $-2\text{Log}(\text{likelihood})$ of full model and null-model is approximately χ^2 -distributed with k degrees of freedom under the null hypotheses $\beta_1 = \dots = \beta_k = 0$. Additional covariates can be included into the models.

Stepwise Regression Analysis

For both multivariate logistic and Cox regression models, a stepwise procedure is used in order to find submodels including only relevant variables. Two effects are usually achieved by these procedures: Variables (methylation ratios) that are basically unrelated to the dependent

variable (response state or DFS, respectively) are excluded as they do not add relevant information to the model. Out of a set of highly correlated variables, only the one with the the best relation to the dependent variable is retained. Inclusion of both types of variables can lead to numerical instabilities and a loss of power. Moreover, the predictory performance can be low due to overfitting. The applied algorithm aims at minimizing the Akaike information criterion (AIC) which is defined as

$$\text{AIC} = -2 \cdot \text{maximized log-likelihood} + 2 \cdot \text{\#parameters}.$$

The AIC is related to the predictory performance of a model, smaller values promise better performance. Whereas the inclusion of additional variables always improves the model fit and thus increases the likelihood, the second term penalizes the estimation of additional parameters. The best model will present a compromise model with good fit and usually a small or moderate number of variables.

Results

Adjuvant setting

Analysis of the methylation patterns of patient samples treated with Tamoxifen as an adjuvant therapy immediately following surgery (see Figure 1) is shown in the plots according to Figures 3 to 45. For each amplificate, the mean methylation over all oligo-pairs for that amplificate was calculated and the population split into groups according to their mean methylation values, wherein one group was composed of individuals with a methylation score higher than the median and a second group composed of individuals with a methylation score lower than the median.

The results are shown in figures 3 to 9, as Cox model estimated disease-free survival curves. Figures 10 to 34 show the disease free survival curves using the methylation analyses of only single oligonucleotide.

In a further analysis the recurrence of distant metastases only was analysed in figures 35 to 46.

The accuracy of the differentiation between the different groups was further increased by combining multiple oligonucleotides from different genes. Figure 53 shows the combination

of two oligonucleotides each from the genes TBC1D3 and CDK6, and one oligonucleotide from the gene PITX2. the broken lines show the classification of the patients from the sample set by means of the St. Gallen method (the current method of choice for estimating disease free survival) as compared to methylation analysis (unbroken lines), thereby showing the improved effectiveness of methylation analysis over current methods, in particular post 80 months. The St. Gallen method is the most commonly used treatment selection criteria for breast cancer patients. The criteria are revised every two years, and are based upon clinical factors (age, type of cancer, size, metastasis etc.), it is used to divide patients into high risk and low risk cases which follow different rules for therapy.

Metastatic setting

Analysis of the methylation patterns of patient samples treated with Tamoxifen in a metastatic setting (see Figure 2) is shown in the matrices according to Figures 46 to 52). The subjects analysed in this classification had relapsed following an initial treatment, the subsequent metastasis being treated by Tamoxifen.

In order to determine the ability of each gene promoter to predict success or failure of Tamoxifen treatment, the individual CpGs measured were combined per gene using Hotelling's T^2 statistics. Several genes were significantly associated with response to tamoxifen after correcting for multiple comparison with a moderate conservative false discovery rate of 25% (see Figure 52). The genes were ONECUT2, WBP11, CYP2D6, DAG1, ERBB2, S100A2, TFF1, TP53, TMEFF2, ESR1, SYK, RASSF1, PITX2, PSAT1, CGA and PCAF.

Figure 50 shows the uncorrected p-values on a log-scale. P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models. Each individual genomic region of interest is represented as a point, the upper dotted line represents the cut off point for the 25% false discovery rate, the lower dotted line shows the Bonferroni corrected 5% limit.

Figure 51 shows a ranked matrix of the best 11 amplificates of data obtained. P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models. The figure is shown in greyscale, wherein the most significant CpG positions are at the bottom of the matrix with significance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in grey, from light (low proportion of methylation) to dark (high

proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the methylation profile for the different CpGs for one sample. The p-values for the individual CpG positions are shown on the right side. The p-values are the probabilities that the observed distribution occurred by chance in the data set.

Figures 47 through 49 the analysis of a subset of shows the uncorrected p-values on a log-scale. Figure 47 shows the uncorrected p-values on a log-scale. P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models according to Example 1 (metastatic setting) . Each individual genomic region of interest is represented as a point, the upper dotted line represents the cut off point for the 25% false discovery rate, the lower dotted line shows the Bonferroni corrected 5% limit.

Figure 48 shows a ranked matrix of the best 11 amplicates of data obtained. P-values were calculated from Likelihood ratio (LR) tests from multivariate logistic regression models. The figure is shown in greyscale, wherein the most significant CpG positions are at the bottom of the matrix with significance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in grey, from light (low proportion of methylation) to dark (high proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the methylation profile for the different CpGs for one sample. The p-values for the individual CpG positions are shown on the right side. The p-values are the probabilities that the observed distribution occurred by chance in the data set.

Real time Quantitative methylation analysis

Genomic DNA was analyzed using the Real Time PCR technique after bisulfite conversion In this analysis four oligonucleotides were used in each reaction. Two non methylation specific PCR primers were used to amplify a segment of the treated genomic DNA containing a methylation variable oligonucleotide probe binding site. Two oligonucleotide probes competitively hybridise to the binding site, one specific for the methylated version of the binding site, the other specific to the unmethylated version of the binding site. Accordingly, one of the probes comprises a CpG at the methylation variable position (i.e. anneals to methylated bisulphite treated sites) and the other comprises a TpG at said position (i.e. anneals to unmethylated bisulphite treated sites). Each species of probe is labelled with a 5' fluorescent reporter dye and a 3' quencher dye wherein the CpG and TpG oligonucleotides are labelled with different dyes.

The reactions are calibrated by reference to DNA standards of known methylation levels in order to quantify the levels of methylation within the sample. The DNA standards were composed of bisulfite treated phi29 amplified genomic DNA (i.e. unmethylated), and/or phi29 amplified genomic DNA treated with Sss1 Methylase enzyme (thereby methylating each CpG position in the sample), which is then treated with bisulfite solution. Seven different reference standards were used with 0%, (i.e. phi29 amplified genomic DNA only), 5%, 10%, 25%, 50%, 75% and 100% (i.e. phi29 Sss1 treated genomic only).

The amount of sample DNA amplified is quantified by reference to the gene (β -actin (*ACTB*)) to normalize for input DNA. For standardization the primers and the probe for analysis of the *ACTB* gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. As there are no methylation variable positions, only one probe oligonucleotide is required.

The following oligonucleotides were used in the reaction:

Primer: TGGTGATGGAGGAGGTTTAGTAAGT (SEQ ID NO: 1088)

Primer: AACCAATAAAACCTACTCCTCCCTTAA (SEQ ID NO: 1089)

Probe: 6FAM-ACCACCACCCAACACACAATAACAAACACA-TAMRA or Dabcyl (SEQ ID NO: 1090)

The extent of methylation at a specific locus was determined by the following formula:

$$\text{methylation rate} = 100 * I_{CG} / (I_{CG} + I_{TG})$$

(I = Intensity of the fluorescence of CG-probe or TG-probe)

Gene PITX2

Primers:

PITX2R02: GTAGGGGAGGGAAGTAGATGTT (SEQ ID NO: 1091)

PITX2Q02: TTCTAATCCTCCTTTCCACAATAA (SEQ ID NO: 1092)

Amplificate length : 143 bp

Probes:

PITX2cg1: FAM-AGTCGGAGTCGGGAGAGCGA-Darquencher (SEQ ID NO: 1093)

PITX2tg1: YAKIMA YELLOW-AGTTGGAGTTGGGAGAGTGAAAGGAGA-Darquencher (SEQ ID NO: 1094)

PCR components: 3 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 62 °C, 1 min

Figure 54 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis. The lower plot shows the proportion of disease free patients in the population with above median methylation levels, the upper plot shows the proportion of disease free patients in the population with below median methylation levels. The X axis shows the disease free survival times of the patients in months, and the Y- axis shows the proportion of disease free survival patients. The p-value (probability that the observed distribution occurred by chance) was calculated as 0.0031, thereby confirming the data obtained by means of array analysis according to figure 6.

Example 2

In order to validate the most promising markers from the microarray study of Example 1 Real-Time assays were designed and optimised in order to provide assays of optimum accuracy. The assays were run on a combination of paraffin embedded tissue (hereinafter also referred to as PET) and fresh frozen tissue samples. DNA derived from PET is often of 'lower quality' (e.g. higher degree of DNA fragmentation and low DNA yield from samples), thus confirmation of assay results on PET demonstrates the robustness of the assay and increased utility of the marker.

Quantitative methylation assays were designed for the genes ERBB2 (SEQ ID NO: 5), TFF1 (SEQ ID NO: 12), PLAU (SEQ ID NO:16), PITX2 (SEQ ID NO:23), ONECUT2 (SEQ ID NO:35), TBC1D3 (SEQ ID NO: 43), and ABCA8 (SEQ ID NO: 49) and tested using a sample set of 415 estrogen receptor positive node negative samples untreated breast cancer patients and 541 estrogen receptor positive node negative samples Tamoxifen treated samples. Approximately 100 of these samples were previously analysed in the microarray study.

The QM assay (= Quantitative Methylation Assay) is a Real-time PCR based method for quantitative DNA methylation detection. The assay principle is based on non-methylation specific amplification of the target region and a methylation specific detection by competitive hybridization of two different probes specific for the CG or the TG status, respectively. For

the present study, TaqMan probes were used that were labeled with two different fluorescence dyes ("FAM" for CG specific probes, "VIC" for TG specific probes) and were further modified by a quencher molecule ("TAMRA" or "Minor Groove Binder/non-fluorescent quencher").

Evaluation of the QM assay raw data is possible with two different methods:

1. Measuring absolute fluorescence intensities (FI) in the logarithmic phase of amplification
2. Difference in threshold cycles (Ct) of CG and TG specific probe.

Results of this study were generated by using the Ct method.

In the following series of quantitative methylation assays the amount of sample DNA amplified is quantified by reference to the gene GSTP1 to normalize for input DNA. For standardization, the primers and the probe for analysis of the GSTP1 gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. As there are no methylation variable positions, only one probe oligonucleotide is required.

Sample Sets

ER+ N0 Untreated Population

To demonstrate that the markers identified have a strong prognostic component, ER+ N0 tumor samples from patients not treated with any adjuvant therapy were analyzed. Markers that are able to show a significant survival difference in this population are considered to be prognostic. All 508 samples of this set were obtained from an academic collaborator as cell nuclei pellets (fresh frozen samples). The sample population can be divided into two subsets: One with 415 randomly selected samples (from both censored and relapsing patients), representing a population with a natural distribution of relapses, and additional 93 samples from relapsing patients only. The latter samples were used for sensitivity/specificity analyses only.

Figure 98 shows the disease-free survival of the randomly selected population in a Kaplan-Meier plot and Figure 99 the distribution of follow-up times for the relapsed and censored patients in histograms. Table 4 lists the number of events broken down by different kinds of relapse. In summary, the survival of this population is comparable to the expected one from the literature.

ER+ N0 TAM treated Population

One intended target population of the invention is patients with ER+ N0 tumors that are treated with hormone therapy. To check the performance of the marker candidates in this population, 589 samples from ER+ N0 tumors from patients treated with Tamoxifen were analyzed. All samples were received as Paraffin-embedded tissues (PET). Three to ten 10 μ m sections were provided.

In addition, for 89 PET patient samples matching fresh frozen samples from the same tumor were included into the study as controls. As these samples were already used in phase 1, they allowed for two kinds of concordance studies :

- Chip *versus* QM assay
- Fresh frozen *versus* PET samples

Samples of the ER+, N0, TAM treated population were received from eight different providers. Altogether 589 samples were processed, 48 of which had to be excluded from the study due to various reasons (e.g. two samples from same tumor, samples from patients that did not fulfill inclusion criteria etc.).

Figure 100 shows the disease-free survival of the total population in a Kaplan-Meier plot and Figure 101 the distribution of follow-up times for the relapsed and censored patients in histograms. Table 5 lists the number of events broken down by different kinds of relapse.

In summary, the survival of this population (82.1 % after 10 years) is comparable to the expected one from the literature (79.2 %).

DNA Extraction

DNA extraction from Fresh Frozen Samples

From a total of 508 fresh frozen samples available as cell nuclei pellets, genomic DNA was isolated using the QIAamp Kit (Qiagen, Hilden, Germany). The extraction was done according to the Cell Culture protocol using Proteinase K with few modifications.

DNA extraction from PET Samples

589 provided PET samples were deparaffinated directly in the tube in which they were delivered by the providers. The tissue was then lysed and DNA extracted using the QIAGEN DNeasy Tissue kit.

Bisulfite treatment

Bisulfite treatment was carried out based on the method disclosed by Olek et al. Nucleic Acids Res. 1996 Dec 15;24(24):5064-6, and optimised to the applicant's laboratory workflow.

Quantification Standards

The reactions are calibrated by reference to DNA standards of known methylation levels in order to quantify the levels of methylation within the sample. The DNA standards were composed of bisulfite treated phi29 amplified human genomic DNA (Promega) (i.e. unmethylated), and/or phi29 amplified genomic DNA treated with Sss1 Methylase enzyme (thereby methylating each CpG position in the sample), which is then treated with bisulfite solution. Seven different reference standards were used with 0%, (i.e. phi29 amplified genomic DNA only), 5%, 10%, 25%, 50%, 75% and 100% (i.e. phi29 Sss1 treated genomic only).

2000 ng batches of human genomic DNA (Promega) were treated with bisulfite. To generate methylated MDA DNA, 13 tubes of 4.5 µg MDA-DNA (700ng/µl) was treated with Sss1.

Control assay

The GSTP1-C3 assay design makes it suitable for quantitating DNAs from different sources, including fresh/frozen samples, remote samples such as plasma or serum, and DNA obtained from archival specimen such as paraffin embedded material.

The following oligonucleotides were used in the reaction to amplify the control amplicate:

Control Primer1: GGAGTGGAGGAAATTGAGAT (SEQ ID NO: 1095)

Control Primer2: CCACACAACAAATACTCAAAAC (SEQ ID NO: 1096)

Control Probe: FAM-TGGGTGTTTGTAATTTTGTGTTTAGGTT-TAMRA (SEQ ID NO: 1097)

Cycle program (40 cycles): 95 °C, 10 min; 95 °C, 15 sec; 58 °C, 1 min

Assay design and reaction conditions

Two assays were developed for the analysis of the gene PITX2(SEQ ID NO: 23)

Assay 1:

Primers: GTAGGGGAGGGAAGTAGATGTT (SEQ ID NO:1098)

TTCTAATCCTCCTTTCCACAATAA (SEQ ID NO:1099)

Probes: FAM-AGTCGGAGTCGGGAGAGCGA-TAMRA (SEQ ID NO:1100)
VIC-AGTTGGAGTTGGGAGAGTGAAAGGAGA -TAMRA (SEQ ID NO:1101)

Amplicon:

GtAGGGGAGGGAAGtAGATGttAGGGtAAGAGTGGAGtGAGtGGAG-
GAGAAAGGAGAGGGGAtTGGGGGtAtTTAGGAGtAAtAGGAGtAGGAG-
tAGAtTttAtTGTGGAAAGGAGGAtAGAA

Length of fragment: 143 bp

Positions of primers, probes and CpG dinucleotides are highlighted.

PCR components (supplied by Eurogentec) : 3 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Cycle program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 62 °C, 1 min

Assay 2 :

Primers: AACATCTACTTCCCTCCCCTAC (SEQ ID NO: 1102)
GTTAGTAGAGATTTTATTAAATTTTATTGTAT (SEQ ID NO: 1103)
Probes: FAM-TTCGGTTGCGCGGT-MGBNQF (SEQ ID NO: 1104)
VIC-TTTGGTTGTGTGGTTG- MGBNQF (SEQ ID NO: 1105)

Amplicon:

GTtAGtAGAGATTtAttAAAtTtAtTGtAtAGTGGGGtGtGtAGt-
tGtTGtGtTGGATtAGGAGAGtAtAGttGGAGtGGGG-
GAGAGtAGGGGAAGAAAAGGtAGGGGAGGGAAGtAGATGtt

Length of fragment: 164 bp

The positions of probes, primers and CpG positions are highlighted.

The probes cover three co-methylated CpG positions.

PCR components (supplied by Eurogentec): 2,5 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

The extent of methylation at a specific locus was determined by the following formulas:

Using absolute fluorescence intensity: methylation rate= $100 * I(CG) / (I(CG) + I(TG))$

(I = Intensity of the fluorescence of CG-probe or TG-probe)

Using threshold cycle Ct: methylation rate= $100 * CG / (CG + TG) = 100 / (1 + TG/CG) = 100 / (1 + 2^{\Delta ct})$

(assuming PCR efficiency E=2; $\Delta Ct = Ct(\text{methylated}) - Ct(\text{unmethylated})$)

Gene PLAU (SEQ ID NO:16)

Primer: GTTAGGTGTATGGGAGGAAGTA (SEQ ID NO: 1106)

TCCCTCCCCTATCTTACAA (SEQ ID NO: 1107)

Probes: FAM-ACCCGAACCCCGCGTACTTC-TAMRA (SEQ ID NO: 1108)

VIC-ACCCAAACCCACATACTTCCACA-TAMRA (SEQ ID NO: 1109)

Amplicon:

GtTAGGTGtATGGGAGGAAGtA GAGAATTTAtAAGtTtT ATTtTtAGTtA-
GA GtTGTTGGGTtttTtCGtTGGAGATCGCGtTTtttAAAtTTTGTGAG TTG-
GAAGtA GGGGTt GGTtTGAG TtTGtAAGAtAGGGGAGGGA

Length of fragment: 166 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec : 2,5 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

Gene ONECUT2 (SEQ ID NO: 35)

Primer: GTAGGAAGAGGTGTTGAGAAATTAA (SEQ ID NO: 1110)

CCACACAAAAAATTTCTATACTCCT (SEQ ID NO:1111)

Probes: FAM- ACGGGTAGAGGCGCGGGT -TAMRA (SEQ ID NO:1112)

VIC- ATGGGTAGAGGTGTGGGTATATTGTTTTG-TAMRA (SEQ ID NO:1113)

Amplicon:

GtAGGAAGAGGTGtTGAGAAATTAAAAATTtAGGTTAGTTAATGtATttTGt-
 GtTGtAGGtTtTTTGtATTAAGGGtTGATTGTGtTGGAtGG-
 GAGGAtTGGGtGGAGGGGAAGGTAGAGGGGTtAtATTGTTtTG-
 GAGtGtTtTTTGtGtTtTtTAGGtAAGtTGAGGTAtAGttTt-
 TATTGTTtTAGGAGtAtAGAAAtTtTGTGTGG

Length of fragment: 266 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec : 3 mM MgCl₂ buffer, 10x buffer, Hotstart
 TAQ, 200 µM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

Gene ABCA8 (SEQ ID NO: 49)

Primer: GTGAGGTATTGGATTtTAGTTTATTTG (SEQ ID NO: 1114)

CCCTAAATCTCATCCTAAAAACAC (SEQ ID NO: 1115)

Probes: FAM- TGAGGTTTCGGTTTTTAACGGTGG -TAMRA (SEQ ID NO: 1116)

VIC- TGAGGTTTGGTTTTTAATGGTGGGAT -TAMRA (SEQ ID NO:

1117)

Amplicon:

GTGAGGTAtTGGATTtAGttATTTGGttAAGtTtTGTTtGAATtGGTGT-
 GTGGGTtGAGGtTtGTTtTAAAGTGGGAtTGGTGTtTAGATGAAATTTGGG-
 GTTTtTGGGtTTTGGTGGGATGTGTtTtAGGATGAGATTAGGG

Length of fragment: 168 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec : 3 mM MgCl₂ buffer, 10x buffer, Hotstart
 TAQ, 200 µM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 62 °C, 1 min

Gene ERBB2 (SEQ ID NO: 5)

Primer: GGAGGGGGTAGAGTTATTAGTTTT (SEQ ID NO: 1118)

ACTCCCAACTTCACTTTCTCC (SEQ ID NO:1119)

Probes: FAM- TAATTTAGGCGTTTCGGCGTTAGG -TAMRA (SEQ ID NO:1120)

VIC- TAATTTAGGTGTTTTGGTGTAGGAGGGA -TAMRA (SEQ ID NO:1121)

Amplicon:

GGAGGGGGTAGAGTTATTAGTTTTTGTATTTAGGGATTTT AGGAAAAGTGTG
AGAA GTTGTAGGTAATTAGG TTT G TTAGGAGGGA TATTAGGTT
TG AAGAGAGGGAGAAAGTGAAGTTGGGAGT

Length of fragment: 144 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec: 2,5 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 62 °C, 1 min

Gene TFF1 (SEQ ID NO: 12)

Primer: AGTTGGTGATGTTGATTAGAGTT (SEQ ID NO: 1122)

CCCTCCCAATATACAAATAAAAACTA (SEQ ID NO:1123)

Probes: FAM- ACACCGTTCGTAAAA-MGBNFQ (SEQ ID NO:1124)

VIC- ACACCATTCATAAAAT-MGBNFQ (SEQ ID NO:1125)

Amplicon:

AGTTGGTGATGTTGATTAGAGTTTTTGTAGTTTTTAAATGATTTTTTTAATTAATTTT
AAATTTTTAGAAATTTAT TATAAAAAGGTTATATTTTTTTGGAGGGA T ATG
GTATTAGGATAGAAGTATTAGGGGATTTTA AA GTGT T AAATAGTAGT
TTTTATTTGTATATTGGGAGGG

Length of fragment: 189 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec: 2,5 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

Gene TBC1D3 (SEQ ID NO: 43)

Primer: TTTTGTAGTTGGTTTTTATTAGGGTTTT (SEQ ID NO: 1126)

CCAACATATCCACCCACTTACT (SEQ ID NO: 1127)

Probes: FAM- TTTCGACTAATCTCCCGCCGA-TAMRA (SEQ ID NO: 1128)

VIC- TTTCAACTAATCTCCCACCAAATTTACTATCA-TAMRA

(SEQ ID NO: 1129)

Amplicon:

tTTtAGtTGGtTttAttAGGGtTtAGAGtttAAGAttAGtAtGGGtTtT-
GGGAAGttTGGtAGtTtTAAAtTtAAtATGttTtATTTGAtAGtAAATTGGGA-
GATtAGtAAAGAGtAAGTGGGTGGATATGtTGG

Length of fragment: 142 bp

The positions of probes, primers and CpG positions are highlighted.

PCR components were supplied by Eurogentec: 4,5 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ, 200 μM dNTP, 625 nM each primer, 200 nM each probe

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

Each of the designed assays was tested on the following sets of samples:

- Tamoxifen treated patients who relapsed during treatment (all relapses).
- Tamoxifen treated patients who relapsed during treatment with distant metastases only.
- Non-Tamoxifen treated patients who relapsed during treatment (all relapses).
- Non-Tamoxifen treated patients who relapsed during treatment with distant metastases only.

Raw Data Processing

All analyses were based on CT evaluation (evaluation using fluorescence intensities are available upon request). Assuming optimal real-time PCR conditions in the exponential amplification phase, the concentration of methylated DNA (C_{meth}) can be determined by

$$C_{meth} = \frac{100}{1 + 2^{(CT_{CG} - CT_{TG})}} [\%],$$

where

CT_{CG} denotes the threshold cycle of the CG reporter (FAM channel) and

CT_{TG} denotes the threshold cycle of the TG reporter (VIC channel).

The thresholds for the cycles were determined by human experts after a visual inspection of the Amplification Plots [ABI PRISM 7900 HT Sequence Detection System User Guide]. The values for the cycles (CT_{CG} and CT_{TG}) were calculated with these thresholds by the ABI 7900 software. Whenever the amplification curve did not exceed the threshold, the value of the cycle was set to the maximum cycle, i.e. 50.

Statistical Methods

Cox Regression

The relation between disease-free survival times (DFS) (or metastasis free survival, MFS) and covariates are modeled using Cox Proportional Hazard models (Cox and Oates, 1984; Harrel, 2001).

The hazard, i.e. the instantaneous risk of a relapse, is modeled as

$$h(t | x) = h_0(t) \cdot \exp(\beta x) \quad (3)$$

and

$$h(t | x_1, \dots, x_k) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_k x_k) \quad (4)$$

for univariate and multiple regression analyses, respectively, where t is the time measured in months after surgery, $h_0(t)$ is the baseline hazard, x is the vector of covariates (e.g. measurements of the assays) and β is the vector of regression coefficients (parameters of the model). β will be estimated by maximizing the partial likelihood of the Cox proportional hazard model. Likelihood ratio tests are performed to test whether methylation is related to the hazard. The difference between $-2 \log(\text{Likelihood})$ of full model and null-model is approximately χ^2 -distributed with k degrees of freedom under the null hypotheses $\beta_1 = \dots = \beta_k = 0$.

The assumption of proportional hazards were checked by scaled Schoenfeld residuals (Therneau *et al.*, 2000). For the calculation, analysis and diagnostic of the Cox Proportional Hazard Model the R functions `coxph`, `coxph.zph` of the “survival” package were used.

Stepwise Regression Analysis

For multivariate Cox regression models a stepwise procedure (Venables *et al.*, 1999; Harrel, 2001) was used in order to find sub-models including only relevant variables. Two effects are usually achieved by these procedures:

- Variables (methylation rates) that are basically unrelated to the dependent variable (DFS/MFS) are excluded as they do not add relevant information to the model.
- Out of a set of highly correlated variables, only the one with the best relation to the dependent variable is retained.

Inclusion of both types of variables can lead to numerical instabilities and a loss of power. Moreover, the predictive performance can be low due to overfitting.

The applied algorithm aims at minimizing the Akaike information criterion (AIC) which is defined as

$$\text{AIC} = -2 \log(\text{maximized log-likelihood}) + 2 \cdot \text{\#parameters}.$$

The AIC is related to the predictive performance of a model, smaller values promise better performance. Whereas the inclusion of additional variables always improves the model fit and thus increases the likelihood, the second term penalizes the estimation of additional parameters. The best model will present a compromise model with good fit and usually a small or moderate number of variables. Stepwise regression calculation with AIC was done with the R function “step”.

Kaplan-Meier Survival Curves and Log-Rank Tests

Survival curves are estimated from DFS/MFS data using the Kaplan-Meier method (Kaplan and Meier, 1958). Log-rank tests were used to test for differences of two survival curves, e.g. survival in hyper- vs. hypomethylated groups. For a description of this test see (Cox and Oates, 1984). For the Kaplan Meier Analysis the functions “survfit” and “survdiff” of the “survival” package were used.

Independence of markers from other covariates

To check whether our marker panel gives additional and independent information, other relevant clinical factors were included in the cox proportional hazard model and the p-values for the weights for every factor were calculated (Wald-Test) (Therneau *et al.*, 2000). For the analysis of additional factors in the Cox Proportional Hazard model, the R function “coxph” was used.

Correlation Analysis

Pearson and Spearman correlation coefficients are calculated to estimate the concordance between measurements (e.g. methylation in matched fresh frozen and PET samples).

Density Estimation

For numerical variables, kernel density estimation was performed with a gaussian kernel and variable bandwidth. The bandwidth is determined using Silverman’s “rule-of-thumb” (Silverman, 1986). For the calculation of the densities the R function “density” was used.

Analysis of Sensitivity and Specificity

For the analysis of sensitivity and specificity of single assays and marker panels ROCs were calculated. The calculation of the ROCs was done with two methods: The first method is to calculate sensitivity and specificity for a given threshold for the time $T_{Threshold}$. With that

threshold, true positives, false positives, true negatives and false negatives were defined and the values for sensitivity and specificity were calculated for different cutoffs of the model. Patients censored before $T_{Threshold}$ were excluded. The ROCs were calculated for different times $T_{Threshold}$ (3 year, 4 years, ... , 10 years). The second method is to calculate sensitivity and specificity by using the Bayes-formula based on the Kaplan-Meier estimates (Heagerty *et al.*, 2000) for the survival probabilities in the marker positive and marker negative groups for a given time $T_{Threshold}$. The ROCs were calculated for different times $T_{Threshold}$ (3 year, 4 years, ... , 10 years).

k-fold Crossvalidation

For the analysis of model selection and model robustness k-fold crossvalidation (Hastie *et al.*, 2001) was used. The set of observation was split in k chunks by random. In turn, every chunk was used as a test set and the remaining k-1 chunks were used as training set. This procedure was repeated n times.

Population Charts

For the description of the relation between censoring and a covariate Population Charts (Möcks *et al.*, 2002) were used. The baseline of the covariate was calculated including all observations with event. For a given time t, the mean (in case of real variables like age) or the fraction (in case of categorical variables) for all censored patients in the risk set at time t was calculated and added to the baseline value.

Technical Performance

Comparison of Assay Replicates

Each marker was measured in at least three replicates, variability between assay replicates was observed to be higher for PET than for fresh frozen samples.

Concordance Study Fresh Frozen versus PET Samples

Markers analyzed in this study (Example 2) were initially identified on a chip platform (Example 1) using fresh frozen samples. The ER+ N0 untreated population was also analyzed on fresh frozen samples in Example 2. A concordance study should demonstrate that measured methylation ratios are comparable for fresh frozen and PET samples. For this purpose, 89 fresh frozen samples from three different providers already used in the chip study were processed again in parallel with a matching PET sample originating from the same tumor.

Figure 97 shows such a concordance study for marker candidate PITX2 assay 1 as a scatter plot between fresh frozen and PET samples (using the QM assay). The association between the paired samples is 0.81 (Spearman's rho). This analysis is based on n=89 samples.

Results

Evaluation of Single Markers

Each of the eight established QM assays was used to measure the 508 samples from the N0, ER+ untreated patient population (random selection and additional relapses) in three replicates. After filtering of measuring points not fulfilling quality criteria and performing a Cox analyses, Kaplan-Meier survival curves and ROC curves for each single marker were generated.

Two different clinical endpoints were used for analyses:

- Disease-free survival, i.e. using all kinds of relapses (distant metastasis, locoregional relapses, relapses at contralateral breast) as event.
- Metastasis-free survival, i.e. treating only distant metastasis as an event.

For analyzing the ER+, N0, TAM treated population, five marker candidates were analyzed on 541 samples from the N0, ER+ untreated patient population. Assays were measured in three replicates. Three assays that were measured on the untreated population (PITX2 -2, ONECUT, and ABCA8) were not measured due to the limited material that was available for the TAM treated population. These assays were rejected either because they performed bad in the untreated population (ONECUT2 and ABCA8) or in case of PITX2-II it performed significantly worse than the other assay of this marker (PITX2-I). After filtering of measuring points not fulfilling quality criteria Kaplan-Meier survival curves and ROC curves for each single marker were generated.

Two different clinical endpoints were used:

- Disease-free survival, i.e. using all kinds of relapses (distant metastasis, locoregional relapses, relapses at contralateral breast) as event.
- Metastasis-free survival, i.e. treating only distant metastasis as an event.

The Kaplan-Meier estimated disease-free survival or metastasis-free survival curves of each single assay are shown in Figures 55 to 80, and combinations of assays are shown in Figures 81 to 96. The X axis shows the disease free survival times of the patients in years, and the Y-

axis shows the proportion of patients with disease free survival. The black plot shows the proportion of disease free patients in the population with above an optimised cut off point's methylation levels, the grey plot shows the proportion of disease free patients in the population with below an optimised cut off point's methylation levels.

The following p-values (probability that the observed distribution occurred by chance) were calculated when the cut off was optimised. For cut-off optimization, the quantiles of both groups were shifted between 0.2 and 0.8 and the p-value for the separation of the curves was calculated for each quantile. The quantile with the lowest p-value was then the best cut-off. Percentage values refer to the methylation ratios at the cut-off point.

Single gene assays

Tamoxifen treated

TAM treated (all relapses) ERBB2 (SEQ ID NO: 5) (Figure 55) : p-value 0.089; cut off point: 1.3%

TAM treated (distant only) ERBB2 (SEQ ID NO: 5) (Figure 56): p-value 0.084; cut off point: 0.1%

TAM treated (all relapses) TFF1 (SEQ ID NO: 12) (Figure 57): p-value 0.037; cut off point: 50.9%

TAM treated (distant only) TFF1 (SEQ ID NO: 12) (Figure 58): p-value 0.029; cut off point: 52.9%

TAM treated (all relapses) PLAUI (SEQ ID NO:16) (Figure 59): p-value 0.056; cut off point: 4.8%

TAM treated (distant only) PLAUI (SEQ ID NO:16) (Figure 60): p-value 0.065; cut off point: 4.8%

TAM treated (all relapses) PITX2(SEQ ID NO:23) (Figure 61): p-value 0.01; cut off point: 13.1%

TAM treated (distant only) PITX2(SEQ ID NO:23) (Figure 62): p-value 0.0012; cut off point: 14.3%

TAM treated (all relapses) TBC1D3 (SEQ ID NO: 43) (assay II) (Figure 63): p-value 0.28; cut off point: 94.6%

TAM treated (distant only) TBC1D3 (SEQ ID NO: 43) (assay II) (Figure 64): p-value 0.078; cut off point: 97%

Figure 103 shows the ROC plot at different times for marker model PITX2 (Assay 1) alone on ER+N0 TAM treated population. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates, and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.42 quantile) are plotted.

AUC 60 months: 0.6

AUC 72 months: 0.69

AUC 84 months: 0.69

AUC 96 months: 0.67

Figure 104 shows the ROC plot at different times for marker model TFF1 on ER+N0 TAM treated population. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates for different thresholds (= 5, 6, 7, 8 years) and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.78 quantile) are plotted.

AUC 60 months: 0.7

AUC 72 months: 0.65

AUC 84 months: 0.61

AUC 96 months: 0.64

Figure 105 shows the ROC plot at different times for marker model PLAUI on ER+N0 TAM treated population. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all relapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates for different thresholds (= 5, 6, 7, 8 years), and the estimated area under the curve (AUC) is calcu-

lated. Values for median cut off (triangle) and best cut off (diamond, 0.77 quantile) are plotted.

AUC 60 months: 0.6

AUC 72 months: 0.63

AUC 84 months: 0.57

AUC 96 months: 0.6

Non Tamoxifen treated

Non Tamoxifen treated (all relapses) ERBB2 (SEQ ID NO: 5) (Figure 65): p-value 0.21; cut off point: 0%

Non Tamoxifen treated (distant only) ERBB2 (SEQ ID NO: 5) (Figure 66): p-value 0.23; cut off point: 0.6%

Non Tamoxifen treated (all relapses) TFF1 (SEQ ID NO: 12) (Figure 67) : p-value 0.012; cut off point: 49.6%

Non Tamoxifen treated (distant only) TFF1 (SEQ ID NO: 12) (Figure 68): p-value 0.016; cut off point: 45.4%

Non Tamoxifen treated (all relapses) PLAU (SEQ ID NO:16) (Figure 69): p-value 0.011; cut off point: 3.2%

Non Tamoxifen treated (distant only) PLAU (SEQ ID NO:16) (Figure 70): p-value 0.0082; cut off point: 5.5%

Non Tamoxifen treated (all relapses) PITX2(SEQ ID NO:23) (I) (Figure 71): p-value 1.4e-06; cut off point: 35.4%

Non Tamoxifen treated (distant only) PITX2(SEQ ID NO:23) (I) (Figure 72): p-value 1.7 e-05; cut off point: 41.2%

Non Tamoxifen treated (all relapses) PITX2(SEQ ID NO:23) (II) (Figure 73): p-value 0.00026; cut off point: 56.1%

Non Tamoxifen treated (distant only) PITX2(SEQ ID NO:23) (II) (Figure 74): p-value 0.0026; cut off point: 61.9%

Non Tamoxifen treated (all relapses) ONECUT2 (SEQ ID NO:35) (Figure 75): p-value 0.26; cut off point: 0%

Non Tamoxifen treated (distant only) ONECUT2 (SEQ ID NO:35) (Figure 76): p-value 0.77; cut off point: 0%

Non Tamoxifen treated (all relapses) TBC1D3 (SEQ ID NO: 43) (Figure 77): p-value 0.004; cut off point: 98.6%

Non Tamoxifen treated (distant only) TBC1D3 (SEQ ID NO: 43) (Figure 78): p-value 0.00022; cut off point: 98.6%

Non Tamoxifen treated (all relapses) ABCA8 (SEQ ID NO: 49) (Figure 79): p-value 0.0065; cut off point: 60.9%

Non Tamoxifen treated (distant only) ABCA8 (SEQ ID NO: 49) (Figure 80): p-value 0.15; cut off point: 49.2%

Panels

Based on the results of the single marker evaluations, it was decided to build models using the marker candidates PITX2-Assay I, TFF1, and PLAU. All possible combinations of these markers were evaluated

Tamoxifen treated

TAM treated (all relapses) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) (Figure 81): p-value 0.023; cut off point: 0.7 quantile

TAM treated (distant only) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) (Figure 82): p-value 0.00084; cut off point: 0.72 quantile

TAM treated (all relapses) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2(SEQ ID NO:23) (Figure 83): p-value 0.037; cut off point: 0.72 quantile

TAM treated (distant only) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2(SEQ ID NO:23) (Figure 84): p-value 0.0014; cut off point: 0.4 quantile

TAM treated (all relapses) PITX2(SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) (Figure 85): p-value 0.17; cut off point: 0.78 quantile

TAM treated (distant only) PITX2(SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) (Figure 86): p-value 0.0048; cut off point: 0.32 quantile

TAM treated (all relapses) PITX2(SEQ ID NO:23) and PLAU (SEQ ID NO:16) (Figure 87): p-value 0.1; cut off point: 0.74 quantile

TAM treated (distant only) PITX2(SEQ ID NO:23) and PLAU (SEQ ID NO:16) (Figure 88): p-value 0.0081; cut off point: 0.44 quantile

Figure 102 shows the ROC plot at different times for marker model PITX2 (Assay 1) and TFF1 on ER+N0 TAM treated population. Figure A shows the plot at 60 months, figure B shows the plot at 72 months, figure C shows the plot at 84 months and figure D shows the plot at 96 months. Only distant metastasis are defined as events. Sensitivity (proportion of all re-

lapsed patients in poor prognostic group) shown on the X-axis and specificity (proportion of all relapse free patients in good prognostic group) shown on the Y-axis are calculated from KM estimates, and the estimated area under the curve (AUC) is calculated. Values for median cut off (triangle) and best cut off (diamond, 0.32 quantile) are plotted.

AUC 60 months: 0.62

AUC 72 months: 0.67

AUC 84 months: 0.63

AUC 96 months: 0.65

Non Tamoxifen treated

Non Tamoxifen treated (all relapses) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) (Figure 89): p-value 0.0015; cut off point: 0.78 quantile

Non Tamoxifen treated (distant only) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) (Figure 90): p-value 0.003; cut off point: 0.8 quantile

Non Tamoxifen treated (all relapses) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2(SEQ ID NO:23) (Figure 91): p-value 8.9e-07; cut off point: 0.64 quantile

Non Tamoxifen treated (distant only) TFF1 (SEQ ID NO: 12) and PLAU (SEQ ID NO:16) and PITX2(SEQ ID NO:23) (Figure 92): p-value 5.4e-05; cut off point: 0.66 quantile

Non Tamoxifen treated (all relapses) PITX2(SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) (Figure 93): p-value 1.9e-06; cut off point: 0.72 quantile

Non Tamoxifen treated (distant only) PITX2(SEQ ID NO:23) and TFF1 (SEQ ID NO: 12) (Figure 94): p-value 3.5e-05; cut off point: 0.76 quantile

Non Tamoxifen treated (all relapses) PITX2(SEQ ID NO:23) and PLAU (SEQ ID NO:16) (Figure 95): p-value 1.1e-06; cut off point: 0.68 quantile

Non Tamoxifen treated (distant only) PITX2(SEQ ID NO:23) and PLAU (SEQ ID NO:16) (Figure 96): p-value 1.5e-05; cut off point: 0.64 quantile

Robustness of marker models

To evaluate the robustness of the models, a crossvalidation was performed on model marker panel PITX2 (Assay 1) plus TFF1 and marker panel PITX2 (Assay 1) alone, with 200 replicates. The stability of the assignment of one certain patient to the bad or good outcome group is illustrated in Figure 106, the left hand figure shows model marker panel PITX2 (Assay 1) plus TFF1 and the right hand figure shows model marker panel PITX2 (Assay 1) alone. The

plot illustrates in how many crossvalidation replicates each patient get's assigned to group 1 (light grey) or group 2 (dark grey).

Figure 107 illustrates the amino acid sequence of the polypeptide encoded by the gene PITX2.

Figure 108 illustrates the positions of the amplicates sequenced in Example 3. 'A' shows an illustration of the gene with the major exons annotated, 'B' shows annotated mRNA transcript variants and 'C' shows CpG rich regions of the gene. The positions of Amplicates 1 to 11 are shown to the right of the illustrations.

Figure 109 shows the sequencing data of 11 amplicates of the gene PITX2 according to Example 3. Each column of the matrices of columns 'A' and 'B' represent the sequencing data for one amplicate. The amplicate number is shown to the left of the matrices. Each row of a matrix represents a single CpG site within the fragment and each column represents an individual DNA sample. The matrices in the column marked 'A' showed below median mehtylation as measured by QM assays, the matrices in the column marked 'B' showed below median mehtylation as measured by QM assays. The bar on the left represents a scale of the percent methylation, with the degree of methylation represented by the shade of each position within the column from black representing 100% methylation to light grey representing 0% methylation. White positions represented a measurement for which no data was available.

Figure 110 shows a schematic view of mRNA transcript variants of PITX2, as annotated in the on-line Ensembl database.

Example 3: Sequencing of gene PITX2

Sequencing of the gene PITX2 was carried out in order to confirm that co-methylation of CpG positions correlated across all exons. For bisulfite sequencing amplification primers were designed to cover 11 sequences within the gene PITX2, see Figure 108 for further details. Sixteen samples analysed in Example 4 were utilized for amplicon production. Each sample was treated with sodium bisulfite and sequenced. Sequence data was obtained using ABI 3700 sequencing technology. Obtained sequence traces were normalized and percentage methylation calculated using the Applicant's proprietary bisulphite sequence sequencing trace analysis program (See WO 2004/000463 for further information).

Samples

Eight samples displayed hypermethylation and eight samples displayed hypomethylation in analysis using QM assay II as described in example 2.

Amplification

Fragments of interest were amplified using the following conditions

PCR Reaction solution :

Taq 5U/ μ l 0,2

dNTPs 25mM each 0,2

10x buffer 2,5

water 10,1

primer (6,25 μ M) 2

DNA (1ng/ μ l) 10

Cycling conditions:

15min 95°C

30s 95°C

30s 58°C

1:30min 72°C

40 cycles

Sequencing

Only G-rich primers were used for sequencing with one exception: Amplificate Number 2 was sequenced using both forward and reverse primer.

ExoSAP-IT Reaction solution:

4 μ l PCR product + 2 μ l ExoSAP-IT

45min/37°C and 15min/95°C

Cycle sequencing:

1 μ l BigDye v.1.1

1 μ l water

4 μ l Sanger buffer

4 μ l dNTP mix (0,025 mM each)

10µl

+

5 µl Primer (2pmol/µl)

+

6µl ExoSAP-IT product

Cycling

2 min 96°C, 26 cycles a (30 s/96°C, 15s/55°C, 4 min/60°C)

Purification

A 96 well MultiScreen (Millipore) plate was filled with Sephadex G50 (Amersham) using an appropriate admeasure device. 300µl water were added to each well and incubated 3h at 4°C. Water was removed by spinning for 5minutes at 910g. Cycle sequencing product was loaded to the plate and purified by spinning for 5min at 910g. 10µl of formamide was added to each eluate.

Results:

All PCRs yielded a product. Figure 109 provides matrices produced from bisulfite sequencing data analysed by the the applicant's proprietary software (See WO 2004/000463 for further information). Each column of the matrices of columns 'A' and 'B' represent the sequencing data for one amplificate. The amplificate number is shown to the left of the matrices. Each row of a matrix represents a single CpG site within the fragment and each column represents an individual DNA sample. The matrices in the column marked 'A' showed below median mehtylation as measured by QM assays (see example 4), the matrices in the column marked 'B' showed below median mehtylation as measured by QM assays. The bar on the left represents a scale of the percent methylation, with the degree of methylation represented by the shade of each position within the column from black representing 100% methylation to light grey representing 0% methylation. White positions represented a measurement for which no data was available.

Bisulfite sequencing indicated differential methylation of CpG sites between the two selected classes of samples, furthermore co-methylation was observed across the gene. In particular amplicates 4 to 7 showed a high level of differential methylation between the two analysed groups.

Table 1

| Accession no. | Gene name/loci | Genomic SEQ ID NO: | Pretreated methylated sequence (sense) SEQ ID NO: | Pretreated methylated strand (antisense) SEQ ID NO: | Pretreated unmethylated sequence (sense) SEQ ID NO: | Pretreated unmethylated sequence (antisense) SEQ ID NO: |
|---------------|----------------|--------------------|---|---|---|---|
| NM_001965 | EGR4 | 1 | 206 | 207 | 328 | 329 |
| NM_000038 | APC | 2 | 208 | 209 | 330 | 331 |
| NM_000077 | CDKN2A | 3 | 210 | 211 | 332 | 333 |
| NM_004385 | CSPG2 | 4 | 212 | 213 | 334 | 335 |
| NM_004448 | ERBB2 | 5 | 214 | 215 | 336 | 337 |
| NM_005563 | STMN1 | 6 | 216 | 217 | 338 | 339 |
| NM_000455 | STK11 | 7 | 218 | 219 | 340 | 341 |
| NM_001216 | CA9 | 8 | 220 | 221 | 342 | 343 |
| NM_001604 | PAX6 | 9 | 222 | 223 | 344 | 345 |
| NM_006142 | SFN | 10 | 224 | 225 | 346 | 347 |
| NM_005978 | S100A2 | 11 | 226 | 227 | 348 | 349 |
| NM_003225 | TFF1 | 12 | 228 | 229 | 350 | 351 |
| NM_003242 | TGFBR2 | 13 | 230 | 231 | 352 | 353 |
| NM_000546 | TP53 | 14 | 232 | 233 | 354 | 355 |
| NM_005427 | TP73 | 15 | 234 | 235 | 356 | 357 |
| NM_002658 | PLAU | 16 | 236 | 237 | 358 | 359 |
| NM_016192 | TMEFF2 | 17 | 238 | 239 | 360 | 361 |
| NM_000125 | ESR1 | 18 | 240 | 241 | 362 | 363 |
| NM_003177 | SYK | 19 | 242 | 243 | 364 | 365 |
| NM_001540 | HSPB1 | 20 | 244 | 245 | 366 | 367 |
| NM_007182 | RASSF1 | 21 | 246 | 247 | 368 | 369 |
| NM_015641 | TES | 22 | 248 | 249 | 370 | 371 |
| NM_000325 | PITX2 | 23 | 250 | 251 | 372 | 373 |
| NM_000836 | GRIN2D | 24 | 252 | 253 | 374 | 375 |
| NM_021154 | PSAT1 | 25 | 254 | 255 | 376 | 377 |
| NM_000735 | CGA | 26 | 256 | 257 | 378 | 379 |
| NM_000106 | CYP2D6 | 27 | 258 | 259 | 380 | 381 |
| NM_004718 | COX7A2L | 28 | 260 | 261 | 382 | 383 |
| NM_001437 | ESR2 | 29 | 262 | 263 | 384 | 385 |
| NM_002658 | PLAU | 30 | 264 | 265 | 386 | 387 |
| NM_000638 | VTN | 31 | 266 | 267 | 388 | 389 |
| NM_001055 | SULT1A1 | 32 | 268 | 269 | 390 | 391 |
| NM_003884 | PCAF | 33 | 270 | 271 | 392 | 393 |
| NM_006254 | PRKCD | 34 | 272 | 273 | 394 | 395 |
| NM_004852 | ONECUT2 | 35 | 274 | 275 | 396 | 397 |
| NM_001706 | BCL6 | 36 | 276 | 277 | 398 | 399 |
| NM_016312 | WBP11 | 37 | 278 | 279 | 400 | 401 |
| NM_002462 | MX1 | 38 | 280 | 281 | 402 | 403 |
| NM_138433 | MX1 | 39 | 282 | 283 | 404 | 405 |
| NM_000484 | APP | 40 | 284 | 285 | 406 | 407 |
| NM_002552 | ORC4L | 41 | 286 | 287 | 408 | 409 |
| NM_138999 | NETO1 | 42 | 288 | 289 | 410 | 411 |

| | | | | | | |
|-----------|--------------------------------------|------|------|------|------|------|
| NM_032258 | TBC1D3 | 43 | 290 | 291 | 412 | 413 |
| NM_005310 | GRB7 | 44 | 292 | 293 | 414 | 415 |
| NM_000106 | CYP2D6 | 45 | 294 | 295 | 416 | 417 |
| NM_001259 | CDK6 | 46 | 296 | 297 | 418 | 419 |
| | Sequence located within Chr. 1p13.2 | 47 | 298 | 299 | 420 | 421 |
| | Sequence located within Chr. 17q25.1 | 48 | 300 | 301 | 422 | 423 |
| NM_007168 | ABCA8 | 49 | 302 | 303 | 424 | 425 |
| | Sequence located within Chr. 12q14.3 | 50 | 304 | 305 | 426 | 427 |
| | Sequence located within Chr. 8q12.1 | 51 | 306 | 307 | 428 | 429 |
| NM_017490 | MARK2 | 52 | 308 | 309 | 430 | 431 |
| NM_005229 | ELK1 | 53 | 310 | 311 | 432 | 433 |
| | Q8WUT3 | 54 | 312 | 313 | 434 | 435 |
| NM_000737 | CGB | 55 | 314 | 315 | 436 | 437 |
| NM_001728 | BSG | 56 | 316 | 317 | 438 | 439 |
| NM_005881 | BCKDK | 57 | 318 | 319 | 440 | 441 |
| NM_014587 | SOX8 | 58 | 320 | 321 | 442 | 443 |
| NM_004393 | DAG1 | 59 | 322 | 323 | 444 | 445 |
| NM_020210 | SEMA4B | 60 | 324 | 325 | 446 | 447 |
| NM_000125 | ESR1 (exon8) | 61 | 204 | 327 | 448 | 449 |
| NM_000325 | PITX2 | 1130 | 1132 | 1133 | 1136 | 1137 |
| NM_003225 | TFF1 | 1131 | 1134 | 1135 | 1138 | 1139 |

Table 2 Primers and amplicates according to Example 1

| Gene: | Primer: | Amplicate Length: |
|--------------------------|---|-------------------|
| EGR4 (SEQ ID NO: 1) | AGGGGGATTGAGTGTTAAGT (SEQ ID NO: 450) CCCAAACATAAACACAAAAT (SEQ ID NO: 451) | 294 |
| APC (SEQ ID NO: 2) | TCAACTACCATCAACTTCCTT A (SEQ ID NO: 452) AATTATTTTGTAGTGTGTTAGT GGG (SEQ ID NO: 453) | 491 |
| CDKN2A (SEQ ID NO: 3) | GGGGTTGGTTGGTTATTAGA (SEQ ID NO: 454) AACCCTCTACCCACCTAAAT (SEQ ID NO: 455) | 256 |
| CSPG2 (SEQ ID NO: 4) | GGATAGGAGTTGGGATTAAG AT | 414 |

| | | |
|---------------------------|---|-----|
| | (SEQ ID NO: 456) AAATCTTTTCAACACCAAAA T (SEQ ID NO: 457) | |
| ERBB2 (SEQ ID NO: 5) | GGAGGGGGTAGAGTTATTAG TT (SEQ ID NO: 458) TATACTTCCTCAAACAACCCT C (SEQ ID NO: 459) | 257 |
| STMN1 (SEQ ID NO: 6) | GAGTTTGTATTTAAGTTGAGT GGTT (SEQ ID NO: 460) AACAAAACAATACCCCTTCTA A (SEQ ID NO: 461) | 334 |
| STMN1 (SEQ ID NO: 6) | CCTCTTACTAACCTCAACCAA C (SEQ ID NO: 463) GAAAGGTAGGGAAGGATTTT T (SEQ ID NO: 462) | 454 |
| STK11 (SEQ ID NO: 7) | TAAAAGAAGGATTTTTGATTG G (SEQ ID NO: 464) CATCTTATTACCTCCCTCCC (SEQ ID NO: 465) | 528 |
| CA9 (SEQ ID NO: 8) | GGGAAGTAGGTTAGGGTTAG TT (SEQ ID NO: 466) AAATCCTCCTCTCAAATAAA T (SEQ ID NO: 467) | |
| PAX6 (SEQ ID NO: 9) | GGAGGGGAGAGGGTTATG (SEQ ID NO: 468) TACTATACACACCCCAAACA A (SEQ ID NO: 469) | 374 |
| SFN (SEQ ID NO: 10) | GAAGAGAGGAGAGGGAGGTA (SEQ ID NO: 470) CTATCCAACAAACCCAACA (SEQ ID NO: 471) | 489 |
| S100A2 (SEQ ID NO: 11) | GTTTTTAAGTTGGAGAAGAGG A (SEQ ID NO: 472) ACCTATAAATCACAACCCACT C (SEQ ID NO: 473) | 460 |
| TFF1 (SEQ ID NO: 12) | TTGGTGATGTTGATTAGAGTT T (SEQ ID NO: 474) | 449 |

| | | |
|---------------------------|---|-----|
| | TAAAACACCTTACATTTTCCC T (SEQ ID NO: 475) | |
| TGFBR2 (SEQ ID NO: 13) | GTAATTTGAAGAAAGTTGAG GG (SEQ ID NO: 476) CCAACAACCTAAACAAAACCT CT (SEQ ID NO: 477) | 296 |
| TP53 (SEQ ID NO: 14) | TTGATGAGAAGAAAGGATTT AGT (SEQ ID NO: 478) TCAAATTCAATCAAAAACCTA CC (SEQ ID NO: 479) | 496 |
| TP73 (SEQ ID NO: 15) | AGTAAATAGTGGGTGAGTTAT GAA (SEQ ID NO: 480) GAAAAACCTCTAAAAACTACT CTCC (SEQ ID NO: 481) | 607 |
| PLAU (SEQ ID NO: 16) | GAGAGAGATAGTTGGGGAGT TT (SEQ ID NO: 482) CAAACAACTTCATCTACCAA ATAC (SEQ ID NO: 483) | 453 |
| TMEFF2 (SEQ ID NO: 17) | TGTTGGTTGTTGTTGTTGTT (SEQ ID NO: 484) CTTTCTACCCATCCCAAAA (SEQ ID NO: 485) | 319 |
| ESR1 (SEQ ID NO: 18) | CTATCAATTCCCCCAACTACT (SEQ ID NO: 487) TTGTTGGATAGAGGTTGAGTT T (SEQ ID NO: 486) | 349 |
| SYK (SEQ ID NO: 19) | GTGGGTTTTGGGTAGTTATAG A (SEQ ID NO: 488) TAACCTCCTCTCCTTACCAA (SEQ ID NO: 489) | 485 |
| HSPB1 (SEQ ID NO: 20) | CCTACCTCTACCACTTCTCAA T (SEQ ID NO: 491) AAGAGGGTTTAGTTTTTATTT GG (SEQ ID NO: 490) | 216 |
| RASSF1 (SEQ ID NO: 21) | AGTGGGTAGGTTAAGTGTGTT G (SEQ ID NO: 492) CCCCAAAATCCAAACTAAA | 319 |

| | | |
|---------------------------|--|-----|
| | (SEQ ID NO: 493) | |
| TES (SEQ ID NO: 22) | AGGTTGGGGATTTTAGTTTTT (SEQ ID NO: 494) ACCTTCTTCACTTTATTTTCCA (SEQ ID NO: 495) | 448 |
| PITX2 (SEQ ID NO: 23) | TCCTCAACTCTACAAACCTAA AA (SEQ ID NO: 497) GTAGGGGAGGGAAGTAGATG T (SEQ ID NO: 496) | 408 |
| GRIN2D (SEQ ID NO: 24) | ATAGTTTGTGGTTTGGATTTTT (SEQ ID NO: 498) AAAACCTTCCCTAACTTCAA T (SEQ ID NO: 499) | 435 |
| PSAT1 (SEQ ID NO: 25) | GTAGGTGGTTAATTTTGGGTT (SEQ ID NO: 500) CTCATTCACACTATATCCATT CA (SEQ ID NO: 501) | 500 |
| PSAT1 (SEQ ID NO: 25) | TAAGAGAGAGGAGTTGAGGT TT (SEQ ID NO: 502) CCAAAATTAACCACCTACCTA A (SEQ ID NO: 503) | 478 |
| CGA (SEQ ID NO: 26) | TAGTGGTATAAGTTTGGAAAT GTT (SEQ ID NO: 504) TCCACCTACATCTAAACCCTA A (SEQ ID NO: 505) | 364 |
| CYP2D6 (SEQ ID NO: 27) | CCTCCTAAACTAAATCCAACA A (SEQ ID NO: 507) GGGGTTAAGGTTTTTATGGTA (SEQ ID NO: 506) | 418 |
| COX7A2L (SEQ ID NO: 28) | AATCCTAAAAACCCTAACTTT TAAT (SEQ ID NO: 509) GGAGGTGTAAGGAGAATAGA GA (SEQ ID NO: 508) | 398 |
| ESR2 (SEQ ID NO: 29) | AAACCTTCCCAATAACCTCTT A (SEQ ID NO: 511) TAGAGGGGAGTAGTGTTTGA GT (SEQ ID NO: 510) | 471 |
| PLAU (SEQ ID NO: 30) | GTGATATTTGGGGATTGTTAT T (SEQ ID NO: 512) ACTCCCTCCCCTATCTTACA (SEQ ID NO: 513) 479 | 479 |
| VTN (SEQ ID NO: 31) | GTTATTTGGGTTAATGTAGGG A (SEQ ID NO: 514) | 492 |

| | | |
|-------------------------|--|-----|
| | TCTATCCCCTCAAACCTTAAAA A (SEQ ID NO: 515) | |
| SULT1A1 (SEQ ID NO: 32) | ATACTACCAAACCCACTCAAA C (SEQ ID NO: 517) GAATTTAGGGAAGGAGTTAG TTG (SEQ ID NO: 516) | 448 |
| PCAF (SEQ ID NO: 33) | GGATAAATGATTGAGAGGTT GT (SEQ ID NO: 518) CCTCCCTTAATTCTCCTACC (SEQ ID NO: 519) | 369 |
| PRKCD (SEQ ID NO: 34) | CTTAACCCATCCCAATCA (SEQ ID NO: 521) GATAGAAGGATTTTAGTTTTT ATTGTT (SEQ ID NO: 520) | 322 |
| ONECUT2 (SEQ ID NO: 35) | TTTGTTGGGATTTGTTAGGAT (SEQ ID NO: 522) AAACATTTTACCCCTCTAAAC C (SEQ ID NO: 523) | 467 |
| BCL6 (SEQ ID NO: 36) | CATCACCCTTCTAAAAACCC (SEQ ID NO: 525) GGGTAAGAAAGAAGGAATTA GTTT (SEQ ID NO: 524) | 456 |
| WBP11 (SEQ ID NO: 37) | AAGAGGTGAGGAAGAGTAGT AAAT (SEQ ID NO: 526) CTCCCAACAATAAATCAAAA T (SEQ ID NO: 527) | 437 |
| MX1 (SEQ ID NO: 38) | TGTAGGAGAGGTTGGGAAG (SEQ ID NO: 528) CCAAACATAACATCCACTAAA A (SEQ ID NO: 529) | 341 |
| MX1 (SEQ ID NO: 39) | TAGGTTTAAGAGGAGAGGGA AT (SEQ ID NO: 530) AAACAACCTACCCAAATCCAA C (SEQ ID NO: 531) | 433 |
| APP (SEQ ID NO: 40) | GAGTAAGGAAGGGGGATG (SEQ ID NO: 532) AACCCAAATCTTTAATACAAA AA (SEQ ID NO: 533) | 494 |
| NETO1 (SEQ ID NO: 42) | GGAGTTTTTTAGAAGAGGAAG ATT (SEQ ID NO: 534) ACTTCACAATAAATACCCTCC C (SEQ ID NO: 535) | 395 |
| TBC1D3 (SEQ ID NO: 43) | GGTAGAGGAAGTAGTTGGTTT G (SEQ ID NO: 536) CTTTTATATTCTCCCAATCTC C (SEQ ID NO: 537) | 490 |
| GRB7 (SEQ ID NO: 44) | AAAATCCATAACCACCAAAA TA (SEQ ID NO: 539) TTAGGAAGTTTTAGGAATGAG G (SEQ ID NO: 538) | 416 |
| CYP2D6 (SEQ ID NO: 45) | AATTTCCTAACCCACTATCCT | 379 |

| | | |
|------------------------|---|-----|
| | C (SEQ ID NO: 541) ATTTGTAGTTTGGGGTGATTT (SEQ ID NO: 540) | |
| CDK6 (SEQ ID NO: 46) | ACCTTAAACACCTTCCCATAA (SEQ ID NO: 543) GTGTAATGATTTTGGATTGAG A (SEQ ID NO: 542) | 456 |
| SEQ ID NO: 47 | AAGGAAGGTAGAGGGTTGAG T (SEQ ID NO: 544) AAAATCCAAAATTAACACCAT T (SEQ ID NO: 545) | 499 |
| SEQ ID NO: 48 | AGTAGATGAAGTTGGGGATT AG (SEQ ID NO: 546) TCCTACTATCCCTTCTCAAAA A (SEQ ID NO: 547) | 500 |
| ABCA8 (SEQ ID NO: 49) | TGATTGTGTAGATTATTTTG GTT (SEQ ID NO: 548) CAAACCTCTCTAAACCTCAATC TC (SEQ ID NO: 549) | 499 |
| SEQ ID NO: 50 | ACCCTAACATTCTCTAAACAA CA (SEQ ID NO: 551) GATGAAAGTGGAAAGATTAT GG (SEQ ID NO: 550) | 441 |
| SEQ ID NO: 51 | CTCCAACCTCTCCTCACCTC (SEQ ID NO: 553) ATTTGAAGGTTGTGTTTGTAG A (SEQ ID NO: 552) | 343 |
| MARK2 (SEQ ID NO: 52) | TCACCACTATCCTCAATAATC A (SEQ ID NO: 555) TAAAGTAGGAAGGTTTGGTTT G (SEQ ID NO: 554) | 476 |
| ELK1 (SEQ ID NO: 53) | CCTCTAATTCCTATCAATCAC C (SEQ ID NO: 557) TTAGAAGTGAAAGTAGAAGG GTTT (SEQ ID NO: 556) | 435 |
| Q8WUT3 (SEQ ID NO: 54) | GGTTAGAAGTTAGAGGGGTA GG (SEQ ID NO: 558) CCATCCCATTACCTATAAAAA T (SEQ ID NO: 559) | 406 |
| CGB (SEQ ID NO: 55) | TCCACCCTATTTTCTACCAA (SEQ ID NO: 561) TTTGTTTTAGGTGGTGTGTAA T (SEQ ID NO: 560) | 417 |
| BSG (SEQ ID NO: 56) | TTATCTATCCCCACACCCTAA T (SEQ ID NO: 563) GGAGTAGGTGAGGAGTATTTT G (SEQ ID NO: 562) | 420 |
| BCKDK (SEQ ID NO: 57) | TCACCTCCTTTTACAACCAAT (SEQ ID NO: 565) TTTGGGAGAGTTTGTAGGATTT A (SEQ ID NO: 564) | 258 |

| | | |
|------------------------------|--|-----|
| SOX8 (SEQ ID NO: 58) | GGGTGGGTAGTAGGTTTGTT (SEQ ID NO: 566) ACACACTCCTTAAACTCTTC C (SEQ ID NO: 567) | 435 |
| DAG1 (SEQ ID NO: 59) | AATACCAACCCAAACATCTAC C (SEQ ID NO: 569) TTTGGTTATGTGGAGTTTATT GT (SEQ ID NO: 568) | 315 |
| ORC4L (SEQ ID NO: 41) | CACTCAAACTTCCCTACCTA C (SEQ ID NO: 571) GGTAATGGTGGGGGTAAAT (SEQ ID NO: 570) | 489 |
| SEMA4B (SEQ ID NO: 60) | ACCAAAATACTACTCCCAAAT C (SEQ ID NO: 573) GGGTAGAGGGAGGTTATTGTT (SEQ ID NO: 572) | 337 |
| ESR1 (exon8) (SEQ ID NO: 61) | TATGATTTGTTGTTGGAGATG T (SEQ ID NO: 574) CTTAAAATCCCTTTAACTATT CCC (SEQ ID NO: 575) | 388 |

Table 3 Hybridisation oligonucleotides according to Example 1

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------|--------------------------------------|
| ONECUT2 (SEQ ID NO: 35) | TACGTAGTTGCGCGTT (SEQ ID NO: 800) |
| ONECUT2 (SEQ ID NO: 35) | GTATGTAGTTGTGTGTT (SEQ ID NO: 801) |
| ONECUT2 (SEQ ID NO: 35) | TTTGTGCGTACGGAT (SEQ ID NO: 802) |
| ONECUT2 (SEQ ID NO: 35) | TTTTTGTGTGTATGGAT (SEQ ID NO: 803) |
| ONECUT2 (SEQ ID NO: 35) | TTAAGCGGGCGTTGAT (SEQ ID NO: 804) |
| ONECUT2 (SEQ ID NO: 35) | TTAAGTGGGTGTTGAT (SEQ ID NO: 805) |
| ONECUT2 (SEQ ID NO: 35) | TAGAGGCGCGGGTTAT (SEQ ID NO: 806) |
| ONECUT2 (SEQ ID NO: 35) | TAGAGGTGTGGGTTAT (SEQ ID NO: 807) |
| BCL6 (SEQ ID NO: 36) | ATTTCGAAATATGTCGG (SEQ ID NO: 1004) |
| BCL6 (SEQ ID NO: 36) | ATTTTGAAATATGTTGGT (SEQ ID NO: 1005) |
| BCL6 (SEQ ID NO: 36) | ATTCGAGACGTTTTGT (SEQ ID NO: 1006) |
| BCL6 (SEQ ID NO: 36) | TTTGAGATGTTTTGTTTA (SEQ ID NO: 1007) |
| BCL6 (SEQ ID NO: 36) | TTCGAGTTTCGAATCGG (SEQ ID NO: 1008) |
| BCL6 (SEQ ID NO: 36) | TTTGAGTTTTGAATTGGA (SEQ ID NO: 1009) |
| BCL6 (SEQ ID NO: 36) | ATAGCGAAGGCGTCGA (SEQ ID NO: 1010) |
| BCL6 (SEQ ID NO: 36) | TATAGTGAAGGTGTTGA (SEQ ID NO: 1011) |
| WBP11 (SEQ ID NO: 37) | TTACGAGAAGCGGGTA (SEQ ID NO: 946) |
| WBP11 (SEQ ID NO: 37) | ATTATGAGAAGTGGGTA (SEQ ID NO: 947) |
| WBP11 (SEQ ID NO: 37) | AGGGGGCGATTTTCGG (SEQ ID NO: 948) |
| WBP11 (SEQ ID NO: 37) | TAGGGGGTGATTTTGG (SEQ ID NO: 949) |
| WBP11 (SEQ ID NO: 37) | TTAGCGTCGTTTGATT (SEQ ID NO: 950) |
| WBP11 (SEQ ID NO: 37) | TTTAGTGTTGTTTGATT (SEQ ID NO: 951) |
| WBP11 (SEQ ID NO: 37) | AGTTCGTTTTATTGCGT (SEQ ID NO: 952) |
| WBP11 (SEQ ID NO: 37) | GAGTTTGTTTTATTGTGT (SEQ ID NO: 953) |
| MX1 (SEQ ID NO: 38) | AACGCGCGAAAGTAAA (SEQ ID NO: 576) |
| MX1 (SEQ ID NO: 38) | TTGGGAATGTGTGAAA (SEQ ID NO: 577) |

| <i>Gene</i> | <i>Oligo:</i> |
|------------------------|---------------------------------------|
| MX1 (SEQ ID NO: 38) | TTCGAGTTGGGTCGAGA (SEQ ID NO: 578) |
| MX1 (SEQ ID NO: 38) | TTTGAGTTGGGTTGAGA (SEQ ID NO: 579) |
| MX1 (SEQ ID NO: 38) | TATGCGCGGGAAGATT (SEQ ID NO: 580) |
| MX1 (SEQ ID NO: 38) | GTATGTGTGGGAAGAT (SEQ ID NO: 581) |
| MX1 (SEQ ID NO: 38) | ATTTACGGTTGCGCGG (SEQ ID NO: 582) |
| MX1 (SEQ ID NO: 38) | TATGGTTGTGTGGGTTA (SEQ ID NO: 583) |
| MX1 (SEQ ID NO: 39) | AGGCGTTTATAGTCGGT (SEQ ID NO: 584) |
| MX1 (SEQ ID NO: 39) | AGGTGTTTATAGTTGGT (SEQ ID NO: 585) |
| MX1 (SEQ ID NO: 39) | TTTCGAGTTCGGAGTA (SEQ ID NO: 586) |
| MX1 (SEQ ID NO: 39) | TTTTGAGTTTGGAGTAG (SEQ ID NO: 587) |
| MX1 (SEQ ID NO: 39) | TTGTCGGTCGTAGCGG (SEQ ID NO: 588) |
| MX1 (SEQ ID NO: 39) | TTTGTTGGTTGTAGTGG (SEQ ID NO: 589) |
| MX1 (SEQ ID NO: 39) | TTCGTTACGGCGGTAG (SEQ ID NO: 590) |
| MX1 (SEQ ID NO: 39) | AGTTTGTTATGGTGGT (SEQ ID NO: 591) |
| APP (SEQ ID NO: 40) | TGAAACGAGGCGGAGA (SEQ ID NO: 592) |
| APP (SEQ ID NO: 40) | TGAAATGAGGTGGAGA (SEQ ID NO: 593) |
| APP (SEQ ID NO: 40) | GACGTTGCGTTTTTCGG (SEQ ID NO: 594) |
| APP (SEQ ID NO: 40) | GGATGTTGTGTTTTTGG (SEQ ID NO: 595) |
| APP (SEQ ID NO: 40) | TTTTTAGCGGGTTCGGA (SEQ ID NO: 596) |
| APP (SEQ ID NO: 40) | TTTTTAGTGGGTTGGA (SEQ ID NO: 597) |
| APP (SEQ ID NO: 40) | GGACGTTTCGTAAGCGG (SEQ ID NO: 598) |
| APP (SEQ ID NO: 40) | GGATGTTTGTAAGTGG (SEQ ID NO: 599) |
| ORC4L (SEQ ID NO: 41) | TTATACGCGTTGTTTAT (SEQ ID NO: 600) |
| ORC4L (SEQ ID NO: 41) | TGTATTATATGTGTTGTTT (SEQ ID NO: 601) |
| ORC4L (SEQ ID NO: 41) | AGCGTGACGGTTCGAG (SEQ ID NO: 602) |
| ORC4L (SEQ ID NO: 41) | AGTGTGATGGTTTGAG (SEQ ID NO: 603) |
| ORC4L (SEQ ID NO: 41) | ATTAGGCGAGTTTCGT (SEQ ID NO: 604) |
| ORC4L (SEQ ID NO: 41) | TTAGGTGAGTTTTGTTT (SEQ ID NO: 605) |
| NETO1 (SEQ ID NO: 42) | TACGTTTCGGTTTTACGA (SEQ ID NO: 606) |
| NETO1 (SEQ ID NO: 42) | TTATGTTTGGTTTTATGAT (SEQ ID NO: 607) |
| NETO1 (SEQ ID NO: 42) | TTACGTCGGTTTCGAT (SEQ ID NO: 608) |
| NETO1 (SEQ ID NO: 42) | TTTATGTTGGTTTTGATT (SEQ ID NO: 609) |
| NETO1 (SEQ ID NO: 42) | TTTCGGTTTCGGGAAAG (SEQ ID NO: 610) |
| NETO1 (SEQ ID NO: 42) | TTTGGTTTTTGGGAAAGG (SEQ ID NO: 611) |
| NETO1 (SEQ ID NO: 42) | TGTCGTACGTGTTTAT (SEQ ID NO: 612) |
| NETO1 (SEQ ID NO: 42) | AATTTTTGTTGTATGTGT (SEQ ID NO: 613) |
| TBC1D3 (SEQ ID NO: 43) | TATTCGCGGGCGGTTT (SEQ ID NO: 988) |
| TBC1D3 (SEQ ID NO: 43) | TAGTATTTGTGGGTGG (SEQ ID NO: 989) |
| TBC1D3 (SEQ ID NO: 43) | ATTCGGCGGGAGATTA (SEQ ID NO: 990) |
| TBC1D3 (SEQ ID NO: 43) | AGTAAATTTGGTGGGA (SEQ ID NO: 991) |
| TBC1D3 (SEQ ID NO: 43) | AGATTAGTCGAAAGAGT (SEQ ID NO: 992) |
| TBC1D3 (SEQ ID NO: 43) | GAGATTAGTTGAAAGAGT (SEQ ID NO: 993) |
| TBC1D3 (SEQ ID NO: 43) | TATATTTTCGGGGTTTTAA (SEQ ID NO: 994) |
| TBC1D3 (SEQ ID NO: 43) | TATATTTTGGGGTTTTAAA (SEQ ID NO: 995) |
| GRB7 (SEQ ID NO: 44) | ATAGTTTCGTTATTTGTAT (SEQ ID NO: 1062) |
| GRB7 (SEQ ID NO: 44) | GGTATAGTTTTGTTATTTG (SEQ ID NO: 1063) |
| GRB7 (SEQ ID NO: 44) | TTTAGTACGGGGTGTA (SEQ ID NO: 1064) |
| GRB7 (SEQ ID NO: 44) | TTTTAGTATGGGGTGTA (SEQ ID NO: 1065) |
| GRB7 (SEQ ID NO: 44) | GGCGTTATAGTTACGTTT (SEQ ID NO: 1066) |

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------------|---------------------------------------|
| GRB7 (SEQ ID NO: 44) | GGGTGTTATAGTTATGTT (SEQ ID NO: 1067) |
| GRB7 (SEQ ID NO: 44) | TGTTTATCGAAGGTAGA (SEQ ID NO: 1068) |
| GRB7 (SEQ ID NO: 44) | TGTTTATTGAAGGTAGAA (SEQ ID NO: 1069) |
| CYP2D6 (SEQ ID NO: 45) | GAGATCGCGTTTTTCGT (SEQ ID NO: 844) |
| CYP2D6 (SEQ ID NO: 45) | AGAGATTGTGTTTTTGT (SEQ ID NO: 845) |
| CYP2D6 (SEQ ID NO: 45) | ATTCGCGGCGAGGATA (SEQ ID NO: 846) |
| CYP2D6 (SEQ ID NO: 45) | GATTTGTGGTGAGGAT (SEQ ID NO: 847) |
| CYP2D6 (SEQ ID NO: 45) | GTCGTTTCGGGGACGT (SEQ ID NO: 848) |
| CYP2D6 (SEQ ID NO: 45) | GTTGTTTTTGGGGATGTG (SEQ ID NO: 849) |
| CYP2D6 (SEQ ID NO: 45) | TAAGTAGCGTCGATAG (SEQ ID NO: 850) |
| CYP2D6 (SEQ ID NO: 45) | AAGTAGTGTTGATAGGG (SEQ ID NO: 851) |
| CDK6 (SEQ ID NO: 46) | TACGAATGCGTGGCGG (SEQ ID NO: 866) |
| CDK6 (SEQ ID NO: 46) | TATGAATGTGTGGTGGA (SEQ ID NO: 867) |
| CDK6 (SEQ ID NO: 46) | TTTCGGAGTAGGCGAG (SEQ ID NO: 868) |
| CDK6 (SEQ ID NO: 46) | TTTTGGAGTAGGTGAG (SEQ ID NO: 869) |
| CDK6 (SEQ ID NO: 46) | TACGTTAGTTTCGCGG (SEQ ID NO: 870) |
| CDK6 (SEQ ID NO: 46) | TATGTTAGTTTTGTGGG (SEQ ID NO: 871) |
| CDK6 (SEQ ID NO: 46) | ATTGAGACGCGTTTGG (SEQ ID NO: 872) |
| CDK6 (SEQ ID NO: 46) | GAGATGTGTTTGGGTA (SEQ ID NO: 873) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TAAATTCGACGGGTTT (SEQ ID NO: 1054) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | ATTTGATGGGTTTTTGT (SEQ ID NO: 1055) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TTTTCGTTTCGGCGGAG (SEQ ID NO: 1056) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TTTGTTTGGTGAGGTT (SEQ ID NO: 1057) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TTTCGCGTTTATCGTGT (SEQ ID NO: 1058) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TGGTTTGTGTTTATTGT (SEQ ID NO: 1059) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TTTCGCGGTTTCGTAGT (SEQ ID NO: 1060) |
| SEQ ID NO: 47 (SEQ ID NO: 47) | TTTGTTGGTTTGTAGTTTA (SEQ ID NO: 1061) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | TTAGGTCGGGAGGAAA (SEQ ID NO: 614) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | TTAGGTTGGGAGGAAA (SEQ ID NO: 615) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | TTAGACGTGGGGCGAT (SEQ ID NO: 616) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | TTAGATGTGGGGTGAT (SEQ ID NO: 617) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | TAAGGTACGAGCGTGT (SEQ ID NO: 618) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | AAGGTATGAGTGTGTG (SEQ ID NO: 619) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | GTAGAGTACGAGAGATT (SEQ ID NO: 620) |
| SEQ ID NO: 48 (SEQ ID NO: 48) | GGTAGAGTATGAGAGAT (SEQ ID NO: 621) |
| ABCA8 (SEQ ID NO: 49) | ATTTGGTTTTCGAAGTTT (SEQ ID NO: 996) |
| ABCA8 (SEQ ID NO: 49) | TATTTGGTTTTGAAGTTT (SEQ ID NO: 997) |
| ABCA8 (SEQ ID NO: 49) | TTTTCGGAATTCGGGT (SEQ ID NO: 998) |
| ABCA8 (SEQ ID NO: 49) | TTTTGGAATTTGGGTGT (SEQ ID NO: 999) |
| ABCA8 (SEQ ID NO: 49) | TTTCGGTTTTTAACGGT (SEQ ID NO: 1000) |
| ABCA8 (SEQ ID NO: 49) | TTTTGGTTTTTAATGGTG (SEQ ID NO: 1001) |
| ABCA8 (SEQ ID NO: 49) | AAAATTTACGAGGGGA (SEQ ID NO: 1002) |
| ABCA8 (SEQ ID NO: 49) | TTAAAATTTATGAGGGGA (SEQ ID NO: 1003) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | ATGACGATGATTGGCGA (SEQ ID NO: 622) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | GATGATGATTGGTGAGT (SEQ ID NO: 623) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | TTATGACGTTTAATCGT (SEQ ID NO: 624) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | AGTTATGATGTTTAATTGT (SEQ ID NO: 625) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | AATCGAACGTTGGCGT (SEQ ID NO: 626) |
| SEQ ID NO: 50 (SEQ ID NO: 50) | AAATTGAATGTTGGTGT (SEQ ID NO: 627) |

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------------|--------------------------------------|
| SEQ ID NO: 51 (SEQ ID NO: 51) | TATTCGGGGTTTCGCGA (SEQ ID NO: 1070) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | ATTTGGGGTTTGTGAG (SEQ ID NO: 1071) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | TATTGTTACGCGTCGA (SEQ ID NO: 1072) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | ATTGTTATGTGTTGATTT (SEQ ID NO: 1073) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | GACGTGTAGGTCGTAT (SEQ ID NO: 1074) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | GATGTGTAGGTTGTATT (SEQ ID NO: 1075) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | TTCGGGAACGATTTTT (SEQ ID NO: 1076) |
| SEQ ID NO: 51 (SEQ ID NO: 51) | GGGTTTGGGAATGATT (SEQ ID NO: 1077) |
| MARK2 (SEQ ID NO: 52) | ATATTTTCGGGGGAAGT (SEQ ID NO: 628) |
| MARK2 (SEQ ID NO: 52) | TATATTTTGGGGGAAGT (SEQ ID NO: 629) |
| MARK2 (SEQ ID NO: 52) | TTTCGTATTTGTCGGA (SEQ ID NO: 630) |
| MARK2 (SEQ ID NO: 52) | TTTGTATTTGTTGGAGT (SEQ ID NO: 631) |
| MARK2 (SEQ ID NO: 52) | GGTTATATCGTAGGGTA (SEQ ID NO: 632) |
| MARK2 (SEQ ID NO: 52) | GGGTTATATTGTAGGGT (SEQ ID NO: 633) |
| MARK2 (SEQ ID NO: 52) | AGGGGGACGAATTAGG (SEQ ID NO: 634) |
| MARK2 (SEQ ID NO: 52) | GAGGGGGATGAATTAG (SEQ ID NO: 635) |
| ELK1 (SEQ ID NO: 53) | GGTCGGCGTTGATTTTA (SEQ ID NO: 920) |
| ELK1 (SEQ ID NO: 53) | GGTTGGTGTGATTTTA (SEQ ID NO: 921) |
| ELK1 (SEQ ID NO: 53) | GTCGGGATTCGAACGG (SEQ ID NO: 922) |
| ELK1 (SEQ ID NO: 53) | GTTGGGATTTGAATGG (SEQ ID NO: 923) |
| ELK1 (SEQ ID NO: 53) | GTCGGAAGTTTCGGGA (SEQ ID NO: 924) |
| ELK1 (SEQ ID NO: 53) | GTTGGAAGTTTTGGGAT (SEQ ID NO: 925) |
| ELK1 (SEQ ID NO: 53) | ATATCGTAGGGTAGGCGG (SEQ ID NO: 926) |
| ELK1 (SEQ ID NO: 53) | ATATTGTAGGGTAGGTGG (SEQ ID NO: 927) |
| Q8WUT3 (SEQ ID NO: 54) | TAGAACGGCGTGGGAT (SEQ ID NO: 636) |
| Q8WUT3 (SEQ ID NO: 54) | TAGAATGGTGTGGGAT (SEQ ID NO: 637) |
| Q8WUT3 (SEQ ID NO: 54) | GTCGCGATGTAGTTACGT (SEQ ID NO: 638) |
| Q8WUT3 (SEQ ID NO: 54) | GTTGTGATGTAGTTATGT (SEQ ID NO: 639) |
| Q8WUT3 (SEQ ID NO: 54) | TTAGTTTTCGGGATCGG (SEQ ID NO: 640) |
| Q8WUT3 (SEQ ID NO: 54) | TTTAGTTTTTGGGATTGG (SEQ ID NO: 641) |
| Q8WUT3 (SEQ ID NO: 54) | TCGTTTTTTCGGGATA (SEQ ID NO: 642) |
| Q8WUT3 (SEQ ID NO: 54) | TTTGTTTTTTGGGATAAA (SEQ ID NO: 643) |
| CGB (SEQ ID NO: 55) | TTACGTCGTGGTTTTTA (SEQ ID NO: 954) |
| CGB (SEQ ID NO: 55) | TTATGTTGTGGTTTTTAG (SEQ ID NO: 955) |
| CGB (SEQ ID NO: 55) | GGCGTGAATTTTCGTGG (SEQ ID NO: 956) |
| CGB (SEQ ID NO: 55) | GGTGTGAATTTTGTGGT (SEQ ID NO: 957) |
| CGB (SEQ ID NO: 55) | TTTCGAGTTTATTCGGT (SEQ ID NO: 958) |
| CGB (SEQ ID NO: 55) | TTTTGAGTTTATTTGGTT (SEQ ID NO: 959) |
| CGB (SEQ ID NO: 55) | TTATCGCGATGTGCGT (SEQ ID NO: 960) |
| CGB (SEQ ID NO: 55) | ATTATTGTGATGTGTGT (SEQ ID NO: 961) |
| BSG (SEQ ID NO: 56) | TACGGTTCGCGTTGTT (SEQ ID NO: 644) |
| BSG (SEQ ID NO: 56) | GGAGTATGGTTTGTGT (SEQ ID NO: 645) |
| BSG (SEQ ID NO: 56) | GTAAGGTTCGGCGAGA (SEQ ID NO: 646) |
| BSG (SEQ ID NO: 56) | GTAAGGTTTGGTGAGA (SEQ ID NO: 647) |
| BSG (SEQ ID NO: 56) | TTACGTTTTTCGGGAAG (SEQ ID NO: 648) |
| BSG (SEQ ID NO: 56) | TTATGTTTTTGGGAAGG (SEQ ID NO: 649) |
| BSG (SEQ ID NO: 56) | TACGTTTTCGAGGATCGG (SEQ ID NO: 650) |
| BSG (SEQ ID NO: 56) | TATGTTTTTGAGGATTGG (SEQ ID NO: 651) |
| BCKDK (SEQ ID NO: 57) | GGGCGTTAGGCGGATT (SEQ ID NO: 652) |

| <i>Gene</i> | <i>Oligo:</i> |
|------------------------|---------------------------------------|
| BCKDK (SEQ ID NO: 57) | TGGGTGTTAGGTGGAT (SEQ ID NO: 653) |
| BCKDK (SEQ ID NO: 57) | AGAGCGGTTAGCGTAG (SEQ ID NO: 654) |
| BCKDK (SEQ ID NO: 57) | TGAGAGTGGTTAGTGT (SEQ ID NO: 655) |
| BCKDK (SEQ ID NO: 57) | ATAGAGGGCGTGAATT (SEQ ID NO: 656) |
| BCKDK (SEQ ID NO: 57) | AGAGGGTGTGAATTTT (SEQ ID NO: 657) |
| BCKDK (SEQ ID NO: 57) | TAGGATTACGAGGAAA (SEQ ID NO: 658) |
| BCKDK (SEQ ID NO: 57) | AGGATTTATGAGGAAAAT (SEQ ID NO: 659) |
| SOX8 (SEQ ID NO: 58) | TTTTCGGTTCGAAGTA (SEQ ID NO: 660) |
| SOX8 (SEQ ID NO: 58) | TTTTGGTTTGAAGTAGG (SEQ ID NO: 661) |
| SOX8 (SEQ ID NO: 58) | AGGTCGTTTTTATCGA (SEQ ID NO: 662) |
| SOX8 (SEQ ID NO: 58) | AGGTTGTTTTTATTGAGT (SEQ ID NO: 663) |
| SOX8 (SEQ ID NO: 58) | GTAGTTACGGGGCGTT (SEQ ID NO: 664) |
| SOX8 (SEQ ID NO: 58) | GTAGTTATGGGGTGTT (SEQ ID NO: 665) |
| SOX8 (SEQ ID NO: 58) | TGTCGTATAGGCGGTT (SEQ ID NO: 666) |
| SOX8 (SEQ ID NO: 58) | TTGTTGTATAGGTGGTT (SEQ ID NO: 667) |
| DAG1 (SEQ ID NO: 59) | TTTCGTGGCGGAGAAT (SEQ ID NO: 820) |
| DAG1 (SEQ ID NO: 59) | TTTTGTGGTGGAGAAT (SEQ ID NO: 821) |
| DAG1 (SEQ ID NO: 59) | TACGGATATTTTCGGTT (SEQ ID NO: 822) |
| DAG1 (SEQ ID NO: 59) | AATTATGGATATTTTGGTT (SEQ ID NO: 823) |
| DAG1 (SEQ ID NO: 59) | TTACGATTCGTAGGTT (SEQ ID NO: 824) |
| DAG1 (SEQ ID NO: 59) | TATTATTATGATTTGTAGGT (SEQ ID NO: 825) |
| SEMA4B (SEQ ID NO: 60) | AGTTTTGGGCGCGATTT (SEQ ID NO: 668) |
| SEMA4B (SEQ ID NO: 60) | AGTTTTGGGTGTGATTT (SEQ ID NO: 669) |
| SEMA4B (SEQ ID NO: 60) | AGCGAATAGATTGCGGAT (SEQ ID NO: 670) |
| SEMA4B (SEQ ID NO: 60) | AGTGAATAGATTGTGGAT (SEQ ID NO: 671) |
| SEMA4B (SEQ ID NO: 60) | AGCGATTAGATTGCGGAT (SEQ ID NO: 672) |
| SEMA4B (SEQ ID NO: 60) | AGTGATTAGATTGTGGAT (SEQ ID NO: 673) |
| SEMA4B (SEQ ID NO: 60) | TAGGCGTTCGATTTTT (SEQ ID NO: 674) |
| SEMA4B (SEQ ID NO: 60) | GGGTAGGTGTTTGATT (SEQ ID NO: 675) |
| APC (SEQ ID NO: 2) | GGTTTCGTTTAATCGT (SEQ ID NO: 928) |
| APC (SEQ ID NO: 2) | GGGTTTTGTTTAATTGTA (SEQ ID NO: 929) |
| APC (SEQ ID NO: 2) | TTCGTATTTAGCGGAT (SEQ ID NO: 930) |
| APC (SEQ ID NO: 2) | GGTTTGTATTTAGTGGA (SEQ ID NO: 931) |
| APC (SEQ ID NO: 2) | ATCGGCGGGTTTTCGA (SEQ ID NO: 932) |
| APC (SEQ ID NO: 2) | AATTGGTGGGTTTTTGA (SEQ ID NO: 933) |
| APC (SEQ ID NO: 2) | ATTTTCGAGTTCGGTA (SEQ ID NO: 934) |
| APC (SEQ ID NO: 2) | TTTTTGAGTTTGGTAGT (SEQ ID NO: 935) |
| CDKN2A (SEQ ID NO: 3) | GGCGTTGTTTAACGTAT (SEQ ID NO: 676) |
| CDKN2A (SEQ ID NO: 3) | GGGTGTTGTTTAATGTA (SEQ ID NO: 677) |
| CDKN2A (SEQ ID NO: 3) | AACGTATCGAATAGTTACGG (SEQ ID NO: 678) |
| CDKN2A (SEQ ID NO: 3) | AATGTATTGAATAGTTATGG (SEQ ID NO: 679) |
| CDKN2A (SEQ ID NO: 3) | TACGGTCGGAGGTCGA (SEQ ID NO: 680) |
| CDKN2A (SEQ ID NO: 3) | TATGGTTGGAGGTGTA (SEQ ID NO: 681) |
| CSPG2 (SEQ ID NO: 4) | TTCGGTTAGTTTCGTAT (SEQ ID NO: 904) |
| CSPG2 (SEQ ID NO: 4) | TTTTGGTTAGTTTTGTATT (SEQ ID NO: 905) |
| CSPG2 (SEQ ID NO: 4) | TTCGGGTATTACGTTT (SEQ ID NO: 906) |
| CSPG2 (SEQ ID NO: 4) | TTTTGGGTATTATGTTTT (SEQ ID NO: 907) |
| CSPG2 (SEQ ID NO: 4) | TTTAGTCGCGTAGCGT (SEQ ID NO: 908) |
| CSPG2 (SEQ ID NO: 4) | ATTTAGTTGTGTAGTGTT (SEQ ID NO: 909) |

| <i>Gene</i> | <i>Oligo:</i> |
|----------------------|---------------------------------------|
| CSPG2 (SEQ ID NO: 4) | AATTCGCGAGTTTAGA (SEQ ID NO: 910) |
| CSPG2 (SEQ ID NO: 4) | GAAAAAAATTTGTGAGTT (SEQ ID NO: 911) |
| ERBB2 (SEQ ID NO: 5) | TGTGAGAACGGTTGTA (SEQ ID NO: 912) |
| ERBB2 (SEQ ID NO: 5) | TGAGAATGGTTGTAGG (SEQ ID NO: 913) |
| ERBB2 (SEQ ID NO: 5) | TTAGGCGTTTCGGCGT (SEQ ID NO: 914) |
| ERBB2 (SEQ ID NO: 5) | TTTAGGTGTTTTGGTGT (SEQ ID NO: 915) |
| ERBB2 (SEQ ID NO: 5) | TAGGTTTGCGCGAAGA (SEQ ID NO: 916) |
| ERBB2 (SEQ ID NO: 5) | TTTGTGTGAAGAGAGG (SEQ ID NO: 917) |
| ERBB2 (SEQ ID NO: 5) | TAATTATCGGAGAAGGA (SEQ ID NO: 918) |
| ERBB2 (SEQ ID NO: 5) | TAATTATTGGAGAAGGAG (SEQ ID NO: 919) |
| STMN1 (SEQ ID NO: 6) | TTAGGCGGTTTCGGATT (SEQ ID NO: 1012) |
| STMN1 (SEQ ID NO: 6) | TTAGGTGGTTTGGATT (SEQ ID NO: 1013) |
| STMN1 (SEQ ID NO: 6) | TATCGGTTTCGGGAATT (SEQ ID NO: 1014) |
| STMN1 (SEQ ID NO: 6) | TATTGGTTTGGGAATTT (SEQ ID NO: 1015) |
| STMN1 (SEQ ID NO: 6) | TTTCGCGCGGAGGTTA (SEQ ID NO: 1016) |
| STMN1 (SEQ ID NO: 6) | TTTTGTGTGGAGGTTA (SEQ ID NO: 1017) |
| STMN1 (SEQ ID NO: 6) | GGTAAGAACGTATATAGT (SEQ ID NO: 1018) |
| STMN1 (SEQ ID NO: 6) | TGGTAAGAATGTATATAGT (SEQ ID NO: 1019) |
| STMN1 (SEQ ID NO: 6) | TTTCGGTTAATGCGGA (SEQ ID NO: 1020) |
| STMN1 (SEQ ID NO: 6) | TTTTTGGTTAATGTGGA (SEQ ID NO: 1021) |
| STMN1 (SEQ ID NO: 6) | TACGTTTCGCGATTTGT (SEQ ID NO: 1022) |
| STMN1 (SEQ ID NO: 6) | AGGGTTATGTTTGTGA (SEQ ID NO: 1023) |
| STMN1 (SEQ ID NO: 6) | GATACGTCGGTGTCTGG (SEQ ID NO: 1024) |
| STMN1 (SEQ ID NO: 6) | TGATATGTTGGTGTGG (SEQ ID NO: 1025) |
| STMN1 (SEQ ID NO: 6) | TTACGGCGAGATTATT (SEQ ID NO: 1026) |
| STMN1 (SEQ ID NO: 6) | TTTTATGGTGAGATTATT (SEQ ID NO: 1027) |
| STK11 (SEQ ID NO: 7) | ATTAATCGTCGTTCTGG (SEQ ID NO: 880) |
| STK11 (SEQ ID NO: 7) | GATTAATTGTTGTTTGGG (SEQ ID NO: 881) |
| STK11 (SEQ ID NO: 7) | TAATCGTTAGCGGCGG (SEQ ID NO: 882) |
| STK11 (SEQ ID NO: 7) | TTAATTGTTAGTGGTGG (SEQ ID NO: 883) |
| STK11 (SEQ ID NO: 7) | GTCGTTTTTCGCGAGGA (SEQ ID NO: 884) |
| STK11 (SEQ ID NO: 7) | GTTGTTTTTGTGAGGAG (SEQ ID NO: 885) |
| STK11 (SEQ ID NO: 7) | TAATGAGCGCGTTGTA (SEQ ID NO: 886) |
| STK11 (SEQ ID NO: 7) | ATGAGTGTGTTGTATTT (SEQ ID NO: 887) |
| CA9 (SEQ ID NO: 8) | ATGGTTTTCGATAATTTTT (SEQ ID NO: 682) |
| CA9 (SEQ ID NO: 8) | ATGGTTTTTGATAATTTTT (SEQ ID NO: 683) |
| CA9 (SEQ ID NO: 8) | TGTACGTATAGTTCGTA (SEQ ID NO: 684) |
| CA9 (SEQ ID NO: 8) | TTAATGTATGTATAGTTTGT (SEQ ID NO: 685) |
| CA9 (SEQ ID NO: 8) | ATATATCGTGTGTTGGG (SEQ ID NO: 686) |
| CA9 (SEQ ID NO: 8) | ATATATTGTGTGTTGGG (SEQ ID NO: 687) |
| CA9 (SEQ ID NO: 8) | ATAGTTAGTCGTATGGT (SEQ ID NO: 688) |
| CA9 (SEQ ID NO: 8) | ATAGTTAGTTGTATGGTT (SEQ ID NO: 689) |
| PAX6 (SEQ ID NO: 9) | TATTGTTTTCGGTTGTTAG (SEQ ID NO: 690) |
| PAX6 (SEQ ID NO: 9) | TATTGTTTTGGTTGTTAG (SEQ ID NO: 691) |
| PAX6 (SEQ ID NO: 9) | GGCGACGCGGTTAGTT (SEQ ID NO: 692) |
| PAX6 (SEQ ID NO: 9) | GGTGATGTGGTTAGTT (SEQ ID NO: 693) |
| PAX6 (SEQ ID NO: 9) | TAGGTCGCGTAGATTT (SEQ ID NO: 694) |
| PAX6 (SEQ ID NO: 9) | AGTTTAGGTTGTGTAGA (SEQ ID NO: 695) |
| PAX6 (SEQ ID NO: 9) | TAGCGTATTTTTTCGGT (SEQ ID NO: 696) |

| <i>Gene</i> | <i>Oligo:</i> |
|------------------------|---|
| PAX6 (SEQ ID NO: 9) | TAGTGTATTTTTTTGGTTG (SEQ ID NO: 697) |
| SFN (SEQ ID NO: 10) | AGTAGGTCGAACGTTA (SEQ ID NO: 698) |
| SFN (SEQ ID NO: 10) | AGAGTAGGTTGAATGTT (SEQ ID NO: 699) |
| SFN (SEQ ID NO: 10) | TTGCGAAGAGCGAAAT (SEQ ID NO: 700) |
| SFN (SEQ ID NO: 10) | TGTGAAGAGTGAAATTT (SEQ ID NO: 701) |
| SFN (SEQ ID NO: 10) | TTCGAGGTGCGTGAGT (SEQ ID NO: 702) |
| SFN (SEQ ID NO: 10) | TTTGAGGTGTGTGAGTA (SEQ ID NO: 703) |
| SFN (SEQ ID NO: 10) | TGTGCGATATCGTGTT (SEQ ID NO: 704) |
| SFN (SEQ ID NO: 10) | TGTGATATTGTGTTGGG (SEQ ID NO: 705) |
| S100A2 (SEQ ID NO: 11) | TTTAATTGCGGTTGTGTG (SEQ ID NO: 786) |
| S100A2 (SEQ ID NO: 11) | TTTAATTGTGGTTGTGTG (SEQ ID NO: 787) |
| S100A2 (SEQ ID NO: 11) | TATATAGGCGTATGTATG (SEQ ID NO: 788) |
| S100A2 (SEQ ID NO: 11) | TATATAGGTGTATGTATG (SEQ ID NO: 789) |
| S100A2 (SEQ ID NO: 11) | TGTATACGAGTATTGGA (SEQ ID NO: 790) |
| S100A2 (SEQ ID NO: 11) | TATGTATATGAGTATTGGA (SEQ ID NO: 791) |
| S100A2 (SEQ ID NO: 11) | AGTTTTAGCGTGTGTTTA (SEQ ID NO: 792) |
| S100A2 (SEQ ID NO: 11) | AGTTTTAGTGTGTGTTTA (SEQ ID NO: 793) |
| TFF1 (SEQ ID NO: 12) | AGAATTTATCGTATAAAAAG (SEQ ID NO: 794) |
| TFF1 (SEQ ID NO: 12) | AATTTATTGTATAAAAAGGT (SEQ ID NO: 795) |
| TFF1 (SEQ ID NO: 12) | GGACGTCGATGGTATT (SEQ ID NO: 796) |
| TFF1 (SEQ ID NO: 12) | AGGGATGTTGATGGTA (SEQ ID NO: 797) |
| TFF1 (SEQ ID NO: 12) | AACGGTGTCGTCGAAA (SEQ ID NO: 798) |
| TFF1 (SEQ ID NO: 12) | AATGGTGTTGTTGAAAT (SEQ ID NO: 799) |
| TGFBR2 (SEQ ID NO: 13) | AAAACGTGGACGTTTT (SEQ ID NO: 896) |
| TGFBR2 (SEQ ID NO: 13) | GAAAATGTGGATGTTTT (SEQ ID NO: 897) |
| TGFBR2 (SEQ ID NO: 13) | TGAAAGTCGGTTAAAGT (SEQ ID NO: 898) |
| TGFBR2 (SEQ ID NO: 13) | TGAAAGTTGGTTAAAGT (SEQ ID NO: 899) |
| TGFBR2 (SEQ ID NO: 13) | TTGGACGTCGAGGAGA (SEQ ID NO: 900) |
| TGFBR2 (SEQ ID NO: 13) | TTGGATGTTGAGGAGA (SEQ ID NO: 901) |
| TGFBR2 (SEQ ID NO: 13) | TTTTCGGGCGGAGAGA (SEQ ID NO: 902) |
| TGFBR2 (SEQ ID NO: 13) | AAGGTTTTTTGGGTGGA (SEQ ID NO: 903) |
| TP53 (SEQ ID NO: 14) | TATTAGGTCGGCGAGA (SEQ ID NO: 858) |
| TP53 (SEQ ID NO: 14) | AGGTTGGTGAGAATTT (SEQ ID NO: 859) |
| TP53 (SEQ ID NO: 14) | TTCCGGTAGGCGGATTA (SEQ ID NO: 860) |
| TP53 (SEQ ID NO: 14) | TTTTTGGTAGGTGGAT (SEQ ID NO: 861) |
| TP53 (SEQ ID NO: 14) | ATATTTTGC GTTCGGG (SEQ ID NO: 862) |
| TP53 (SEQ ID NO: 14) | ATATTTTGTGTTTGGGT (SEQ ID NO: 863) |
| TP53 (SEQ ID NO: 14) | TACGACGGTGATACGT (SEQ ID NO: 864) |
| TP53 (SEQ ID NO: 14) | TTTATGATGGTGATATGT (SEQ ID NO: 865) |
| TP73 (SEQ ID NO: 15) | TTCGTTTCGCGAAGTTA (SEQ ID NO: 706) |
| TP73 (SEQ ID NO: 15) | GGTTTGTGTTGTGAAGTTA (SEQ ID NO: 707) |
| PLAU (SEQ ID NO: 16) | AAGAGGTCGTCGGGAT (SEQ ID NO: 708) |
| PLAU (SEQ ID NO: 16) | AAGAGGTTGTTGGGAT (SEQ ID NO: 709) |
| PLAU (SEQ ID NO: 16) | TTATCGCGGGTATTTT (SEQ ID NO: 710) |
| PLAU (SEQ ID NO: 16) | TTGGTTATTGTGGGTAT (SEQ ID NO: 711) |
| PLAU (SEQ ID NO: 16) | TTCGATTTCGTTATTATG (SEQ ID NO: 712) |
| PLAU (SEQ ID NO: 16) | TTTGATTTTGTGTTATTATGAG (SEQ ID NO: 713) |
| PLAU (SEQ ID NO: 16) | GTCGTGAGCGATTTTA (SEQ ID NO: 714) |
| PLAU (SEQ ID NO: 16) | TTGGTTGTGAGTGATT (SEQ ID NO: 715) |

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------|---------------------------------------|
| TMEFF2 (SEQ ID NO: 17) | TATCGTAGTTCGTTTCGG (SEQ ID NO: 874) |
| TMEFF2 (SEQ ID NO: 17) | ATTGTAGTTTGTGTTGGT (SEQ ID NO: 875) |
| TMEFF2 (SEQ ID NO: 17) | AAACGTTTATCGGTTG (SEQ ID NO: 876) |
| TMEFF2 (SEQ ID NO: 17) | AATGTTTATTGGTTGGA (SEQ ID NO: 877) |
| TMEFF2 (SEQ ID NO: 17) | TTCGTAGAAGAATACGCGTA (SEQ ID NO: 878) |
| TMEFF2 (SEQ ID NO: 17) | TTGTAGAAGAATATGTGTA (SEQ ID NO: 879) |
| ESR1 (SEQ ID NO: 18) | TGCGGTTGTATACGTAG (SEQ ID NO: 962) |
| ESR1 (SEQ ID NO: 18) | TGTGTGGTTGTATATGT (SEQ ID NO: 963) |
| ESR1 (SEQ ID NO: 18) | TTCGTGTTAGATTTCGATAT (SEQ ID NO: 964) |
| ESR1 (SEQ ID NO: 18) | TTTGTGTTAGATTTTGATAT (SEQ ID NO: 965) |
| ESR1 (SEQ ID NO: 18) | AACGCGAAAGACGGAT (SEQ ID NO: 966) |
| ESR1 (SEQ ID NO: 18) | ATAAATGTGAAAGATGGA (SEQ ID NO: 967) |
| ESR1 (SEQ ID NO: 18) | GGGCGTACGAGGATTT (SEQ ID NO: 968) |
| ESR1 (SEQ ID NO: 18) | GGGTGTATGAGGATTT (SEQ ID NO: 969) |
| HSPB1 (SEQ ID NO: 20) | AGGGTATTCGTCGGTT (SEQ ID NO: 888) |
| HSPB1 (SEQ ID NO: 20) | AGGGTATTTGTTGGTT (SEQ ID NO: 889) |
| HSPB1 (SEQ ID NO: 20) | GAATTCGAGAGCGCGA (SEQ ID NO: 892) |
| HSPB1 (SEQ ID NO: 20) | TGAATTTGAGAGTGTGA (SEQ ID NO: 893) |
| RASSF1 (SEQ ID NO: 21) | AGTAAATCGGATTAGGA (SEQ ID NO: 852) |
| RASSF1 (SEQ ID NO: 21) | AGTAAATTGGATTAGGAG (SEQ ID NO: 853) |
| RASSF1 (SEQ ID NO: 21) | TACGGGTATTTTCGCGT (SEQ ID NO: 854) |
| RASSF1 (SEQ ID NO: 21) | ATATGGGTATTTTGTGT (SEQ ID NO: 855) |
| RASSF1 (SEQ ID NO: 21) | TGCGAGAGCGCGTTTA (SEQ ID NO: 856) |
| RASSF1 (SEQ ID NO: 21) | TTGTGAGAGTGTGTTTA (SEQ ID NO: 857) |
| GRIN2D (SEQ ID NO: 24) | ATTTCGATTGAGGCGG (SEQ ID NO: 716) |
| GRIN2D (SEQ ID NO: 24) | ATTTTGATTGAGGTGG (SEQ ID NO: 717) |
| PSAT1 (SEQ ID NO: 25) | TTCGTCGGTGTTACGT (SEQ ID NO: 718) |
| PSAT1 (SEQ ID NO: 25) | TTTTGTTGGTGTTATGT (SEQ ID NO: 719) |
| PSAT1 (SEQ ID NO: 25) | GGCGAGTTCGGGTAGT (SEQ ID NO: 720) |
| PSAT1 (SEQ ID NO: 25) | GGTGAGTTTGGGTAGT (SEQ ID NO: 721) |
| PSAT1 (SEQ ID NO: 25) | ATAGTAAACGCGAGGA (SEQ ID NO: 818) |
| PSAT1 (SEQ ID NO: 25) | AGTAAATGTGAGGAGG (SEQ ID NO: 819) |
| PSAT1 (SEQ ID NO: 25) | AAGTTTTTCGCGAGCGG (SEQ ID NO: 722) |
| PSAT1 (SEQ ID NO: 25) | AAGTTTTTGTGAGTGG (SEQ ID NO: 723) |
| PSAT1 (SEQ ID NO: 25) | AGGAAGTTCGGCGAGG (SEQ ID NO: 724) |
| PSAT1 (SEQ ID NO: 25) | AGGAAGTTTGGTGAGG (SEQ ID NO: 725) |
| CYP2D6 (SEQ ID NO: 27) | TACGACGATTTTCGTT (SEQ ID NO: 726) |
| CYP2D6 (SEQ ID NO: 27) | GAGTATGATGATTTTGT (SEQ ID NO: 727) |
| CYP2D6 (SEQ ID NO: 27) | TTCGTCGATTAAGTCGG (SEQ ID NO: 728) |
| CYP2D6 (SEQ ID NO: 27) | TTTGTTGATTAAGTTGGT (SEQ ID NO: 729) |
| CYP2D6 (SEQ ID NO: 27) | GTGGCGCGAGTAGAGG (SEQ ID NO: 730) |
| CYP2D6 (SEQ ID NO: 27) | GTGGTGTGAGTAGAGG (SEQ ID NO: 731) |
| CYP2D6 (SEQ ID NO: 27) | AACGTTTACGTGTTTCGT (SEQ ID NO: 732) |
| CYP2D6 (SEQ ID NO: 27) | GTAATGTTTATGTGTTTGT (SEQ ID NO: 733) |
| COX7A2L (SEQ ID NO: 28) | AATTCGATCGCGGGTA (SEQ ID NO: 1086) |
| COX7A2L (SEQ ID NO: 28) | ATTTGATTGTGGGTAGA (SEQ ID NO: 1087) |
| PLAU (SEQ ID NO: 30) | TATTTGTCGCGTTGAT (SEQ ID NO: 1044) |
| PLAU (SEQ ID NO: 30) | ATTTGTTGTGTTGATGA (SEQ ID NO: 1045) |
| PLAU (SEQ ID NO: 30) | TGTAATTCGGGGATTT (SEQ ID NO: 1046) |

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------|--------------------------------------|
| PLAU (SEQ ID NO: 30) | TTGTAATTTGGGGGATTT (SEQ ID NO: 1047) |
| PLAU (SEQ ID NO: 30) | AGGAAGTACGGAGAAT (SEQ ID NO: 1048) |
| PLAU (SEQ ID NO: 30) | AGGAAGTATGGAGAATT (SEQ ID NO: 1049) |
| PLAU (SEQ ID NO: 30) | TTCGTTGGAGATCGCGT (SEQ ID NO: 1050) |
| PLAU (SEQ ID NO: 30) | TTTGTTGGAGATTGTGT (SEQ ID NO: 1051) |
| PLAU (SEQ ID NO: 30) | TTGCGGAAGTACGCGG (SEQ ID NO: 1052) |
| PLAU (SEQ ID NO: 30) | TTGTGGAAGTATGTGG (SEQ ID NO: 1053) |
| VTN (SEQ ID NO: 31) | TTCGGGTTCGCGAAAG (SEQ ID NO: 1028) |
| VTN (SEQ ID NO: 31) | TTTGGGTTTGTGAAAG (SEQ ID NO: 1029) |
| VTN (SEQ ID NO: 31) | TTTTGTTCGCGTTGAA (SEQ ID NO: 1030) |
| VTN (SEQ ID NO: 31) | TTGTTTGTGTTGAAGTA (SEQ ID NO: 1031) |
| VTN (SEQ ID NO: 31) | TGGGTCGCGAGGTAGT (SEQ ID NO: 1032) |
| VTN (SEQ ID NO: 31) | TGGGTTGTGAGGTAGT (SEQ ID NO: 1033) |
| VTN (SEQ ID NO: 31) | TTCGATGGCGGTTTCGA (SEQ ID NO: 1036) |
| VTN (SEQ ID NO: 31) | TTTGATGGTGGTTTTGA (SEQ ID NO: 1037) |
| SULT1A1 (SEQ ID NO: 32) | TTCGAGTCGTTTTGAT (SEQ ID NO: 734) |
| SULT1A1 (SEQ ID NO: 32) | TTTGAGTTGTTTTGATG (SEQ ID NO: 735) |
| SULT1A1 (SEQ ID NO: 32) | TTCGTCGTGTACGGTT (SEQ ID NO: 736) |
| SULT1A1 (SEQ ID NO: 32) | TTTGTTGTGTATGGTTT (SEQ ID NO: 737) |
| SULT1A1 (SEQ ID NO: 32) | AGGATTTTCGTTTTTCGG (SEQ ID NO: 738) |
| SULT1A1 (SEQ ID NO: 32) | AGGATTTTGTTTTTTGGG (SEQ ID NO: 739) |
| SULT1A1 (SEQ ID NO: 32) | TTTTTCGGTTGAAGTCGG (SEQ ID NO: 740) |
| SULT1A1 (SEQ ID NO: 32) | TTTTTGGTTGAAGTTGG (SEQ ID NO: 741) |
| PCAF (SEQ ID NO: 33) | AGCGTCGGTACGTATA (SEQ ID NO: 986) |
| PCAF (SEQ ID NO: 33) | GGTAGTGTTGGTATGT (SEQ ID NO: 987) |
| PRKCD (SEQ ID NO: 34) | ATTTCGCGTTCGGATT (SEQ ID NO: 742) |
| PRKCD (SEQ ID NO: 34) | GATTTTGTGTTTGGATT (SEQ ID NO: 743) |
| EGR4 (SEQ ID NO: 1) | AAGCGTATTTATCGGA (SEQ ID NO: 744) |
| EGR4 (SEQ ID NO: 1) | GGAAGTGTATTTATTGGA (SEQ ID NO: 745) |
| EGR4 (SEQ ID NO: 1) | TATCGGACGGTCGGTT (SEQ ID NO: 746) |
| EGR4 (SEQ ID NO: 1) | ATTTATTGGATGGTTGG (SEQ ID NO: 747) |
| EGR4 (SEQ ID NO: 1) | AGGCGTAGCGTTTTAG (SEQ ID NO: 748) |
| EGR4 (SEQ ID NO: 1) | TGAGGTGTAGTGTTTT (SEQ ID NO: 749) |
| EGR4 (SEQ ID NO: 1) | AACGTTATAGTTCGAGT (SEQ ID NO: 750) |
| EGR4 (SEQ ID NO: 1) | AATGTTATAGTTTGAGTTT (SEQ ID NO: 751) |
| TP73 (SEQ ID NO: 15) | GTGCGAGTTAGTCGGA (SEQ ID NO: 752) |
| TP73 (SEQ ID NO: 15) | GTGTGAGTTAGTTGGA (SEQ ID NO: 753) |
| TP73 (SEQ ID NO: 15) | TATCGGTTTCGGAGTTA (SEQ ID NO: 754) |
| TP73 (SEQ ID NO: 15) | AGGATATTGGTTTGGAG (SEQ ID NO: 755) |
| TP73 (SEQ ID NO: 15) | AGAGTCGTTTCGGAATT (SEQ ID NO: 756) |
| TP73 (SEQ ID NO: 15) | TGAGAGTTGTTTGGAAAT (SEQ ID NO: 757) |
| SYK (SEQ ID NO: 19) | GAAGTTATCGCGTTGG (SEQ ID NO: 826) |
| SYK (SEQ ID NO: 19) | AGAAGTTATTGTGTTGG (SEQ ID NO: 827) |
| SYK (SEQ ID NO: 19) | GATCGATGCGGTTTAT (SEQ ID NO: 828) |
| SYK (SEQ ID NO: 19) | GGGATTGATGTGGTTTA (SEQ ID NO: 829) |
| SYK (SEQ ID NO: 19) | GTTCGGCGGGAGGAGA (SEQ ID NO: 830) |
| SYK (SEQ ID NO: 19) | GTTTGGTGGGAGGAGA (SEQ ID NO: 831) |
| SYK (SEQ ID NO: 19) | AGTCGATTTTCGTTTAG (SEQ ID NO: 832) |
| SYK (SEQ ID NO: 19) | TAGTTGATTTTGTTTAGT (SEQ ID NO: 833) |

| <i>Gene</i> | <i>Oligo:</i> |
|-------------------------|---------------------------------------|
| SYK (SEQ ID NO: 19) | GGAAGAGTCGCGGGT (SEQ ID NO: 834) |
| SYK (SEQ ID NO: 19) | GGAAGAGTTGTGGGT (SEQ ID NO: 835) |
| HSPB1 (SEQ ID NO: 20) | AGTCGTGTTACGGTAG (SEQ ID NO: 890) |
| HSPB1 (SEQ ID NO: 20) | AGTTGTGTTATGGTAGG (SEQ ID NO: 891) |
| HSPB1 (SEQ ID NO: 20) | TTTTTTCGTTAAGGAAAG (SEQ ID NO: 894) |
| HSPB1 (SEQ ID NO: 20) | TTTTTTTGTGTTAAGGAAAG (SEQ ID NO: 895) |
| TES (SEQ ID NO: 22) | TAGAAGTCGGTTCGTG (SEQ ID NO: 758) |
| TES (SEQ ID NO: 22) | AGAAGTTGGTTTGTGG (SEQ ID NO: 759) |
| TES (SEQ ID NO: 22) | GATTGGGCGGCGGAAG (SEQ ID NO: 760) |
| TES (SEQ ID NO: 22) | ATTGGGTGGTGGAAAGT (SEQ ID NO: 761) |
| TES (SEQ ID NO: 22) | TAGCGGAGTCGGAGGT (SEQ ID NO: 762) |
| TES (SEQ ID NO: 22) | TAGTGGAGTTGGAGGT (SEQ ID NO: 763) |
| TES (SEQ ID NO: 22) | AATTCGGTCGTGGGAT (SEQ ID NO: 764) |
| TES (SEQ ID NO: 22) | AATTTGGTTGTGGGAT (SEQ ID NO: 765) |
| PITX2 (SEQ ID NO: 23) | AGTCGGGAGAGCGAAA (SEQ ID NO: 970) |
| PITX2 (SEQ ID NO: 23) | AGTTGGGAGAGTGAAA (SEQ ID NO: 971) |
| PITX2 (SEQ ID NO: 23) | AAGAGTCGGGAGTCGGA (SEQ ID NO: 972) |
| PITX2 (SEQ ID NO: 23) | AAGAGTTGGGAGTTGGA (SEQ ID NO: 973) |
| PITX2 (SEQ ID NO: 23) | GGTCGAAGAGTCGGGA (SEQ ID NO: 974) |
| PITX2 (SEQ ID NO: 23) | GGTTGAAGAGTTGGGA (SEQ ID NO: 975) |
| PITX2 (SEQ ID NO: 23) | ATGTTAGCGGGTCGAA (SEQ ID NO: 976) |
| PITX2 (SEQ ID NO: 23) | TAGTGGGTTGAAGAGT (SEQ ID NO: 977) |
| GRIN2D (SEQ ID NO: 24) | GAGAGTCGGGATGATT (SEQ ID NO: 766) |
| GRIN2D (SEQ ID NO: 24) | GGAGAGTTGGGATGAT (SEQ ID NO: 767) |
| GRIN2D (SEQ ID NO: 24) | TAGGGTCGAGATTTGG (SEQ ID NO: 768) |
| GRIN2D (SEQ ID NO: 24) | TTAGGGTTGAGATTTGG (SEQ ID NO: 769) |
| GRIN2D (SEQ ID NO: 24) | AGTGTGGCGAATATTG (SEQ ID NO: 770) |
| GRIN2D (SEQ ID NO: 24) | GTGTGGTGAATATTGAA (SEQ ID NO: 771) |
| PSAT1 (SEQ ID NO: 25) | TTTCGATTTCGGTTAGA (SEQ ID NO: 808) |
| PSAT1 (SEQ ID NO: 25) | AATTGTTTTGATTTGGTT (SEQ ID NO: 809) |
| PSAT1 (SEQ ID NO: 25) | TAATGGGGCGTCGATT (SEQ ID NO: 810) |
| PSAT1 (SEQ ID NO: 25) | TTAATGGGGTGTTGATT (SEQ ID NO: 811) |
| PSAT1 (SEQ ID NO: 25) | TATCGTAGCGGTTAGG (SEQ ID NO: 812) |
| PSAT1 (SEQ ID NO: 25) | TATTGTAGTGGTTAGGAA (SEQ ID NO: 813) |
| PSAT1 (SEQ ID NO: 25) | AGGAACGTTAGTCGTT (SEQ ID NO: 814) |
| PSAT1 (SEQ ID NO: 25) | TAGGAATGTTAGTTGTTT (SEQ ID NO: 815) |
| PSAT1 (SEQ ID NO: 25) | GGTCGTCGTATTATGGA (SEQ ID NO: 816) |
| PSAT1 (SEQ ID NO: 25) | TGGTTGTTGTATTATGGA (SEQ ID NO: 817) |
| CGA (SEQ ID NO: 26) | ATATTTATTTTCGGAAATT (SEQ ID NO: 836) |
| CGA (SEQ ID NO: 26) | TTATTTTGGAAATTTATAGT (SEQ ID NO: 837) |
| CGA (SEQ ID NO: 26) | TGATTTTGTCTGTTATTATT (SEQ ID NO: 838) |
| CGA (SEQ ID NO: 26) | TTGATTTTGTGTTATTATT (SEQ ID NO: 839) |
| CGA (SEQ ID NO: 26) | TAAATTGACGTTATGGTA (SEQ ID NO: 840) |
| CGA (SEQ ID NO: 26) | AAATTGATGTTATGGTAAA (SEQ ID NO: 841) |
| CGA (SEQ ID NO: 26) | AATTGACGTTATGGTAAT (SEQ ID NO: 842) |
| CGA (SEQ ID NO: 26) | TAAAAATTGATGTTATGGT (SEQ ID NO: 843) |
| COX7A2L (SEQ ID NO: 28) | TTGTTCTGAAGATCGTT (SEQ ID NO: 1078) |
| COX7A2L (SEQ ID NO: 28) | GTTGTTTGAAGATTGTTT (SEQ ID NO: 1079) |
| COX7A2L (SEQ ID NO: 28) | TAGCGTAAGGATTCGGT (SEQ ID NO: 1080) |

| <i>Gene</i> | <i>Oligo:</i> |
|------------------------------|---------------------------------------|
| COX7A2L (SEQ ID NO: 28) | TTAGTGTAAGGATTGTT (SEQ ID NO: 1081) |
| COX7A2L (SEQ ID NO: 28) | AGAGTTCGGTTTTTCGTA (SEQ ID NO: 1082) |
| COX7A2L (SEQ ID NO: 28) | AGAGTTTGGTTTTTTGTA (SEQ ID NO: 1083) |
| COX7A2L (SEQ ID NO: 28) | ATTCGTATTTGCGGGTTA (SEQ ID NO: 1084) |
| COX7A2L (SEQ ID NO: 28) | ATTTGTATTTGTGGGTTA (SEQ ID NO: 1085) |
| ESR2 (SEQ ID NO: 29) | ATTTTCGAGGATTACGTT (SEQ ID NO: 936) |
| ESR2 (SEQ ID NO: 29) | ATTTTGAGGATTATGTTTT (SEQ ID NO: 937) |
| ESR2 (SEQ ID NO: 29) | AGATGGCGTTTTTCGTA (SEQ ID NO: 938) |
| ESR2 (SEQ ID NO: 29) | TAGATGGTGTTTTTTGTA (SEQ ID NO: 939) |
| ESR2 (SEQ ID NO: 29) | ATTTTCGAATCGATTTTT (SEQ ID NO: 940) |
| ESR2 (SEQ ID NO: 29) | GGAGTATTTTTGAATTGAT (SEQ ID NO: 941) |
| ESR2 (SEQ ID NO: 29) | AGTTCGACGGTTTTAG (SEQ ID NO: 942) |
| ESR2 (SEQ ID NO: 29) | AGGGAGTTTGATGGTT (SEQ ID NO: 943) |
| ESR2 (SEQ ID NO: 29) | AGTTTACGTGATCGAG (SEQ ID NO: 944) |
| ESR2 (SEQ ID NO: 29) | AGTTTATGTGATTGAGTT (SEQ ID NO: 945) |
| VTN (SEQ ID NO: 31) | GGTGGTATCGATTGAT (SEQ ID NO: 1034) |
| VTN (SEQ ID NO: 31) | TGGTGGTATTGATTGAT (SEQ ID NO: 1035) |
| VTN (SEQ ID NO: 31) | TAGTGATTCGCGGGGA (SEQ ID NO: 1038) |
| VTN (SEQ ID NO: 31) | TAGTGATTTGTGGGGA (SEQ ID NO: 1039) |
| VTN (SEQ ID NO: 31) | TTATGTCGGAGGATGA (SEQ ID NO: 1040) |
| VTN (SEQ ID NO: 31) | ATTATGTTGGAGGATGA (SEQ ID NO: 1041) |
| VTN (SEQ ID NO: 31) | ATACGGTTTATGACGAT (SEQ ID NO: 1042) |
| VTN (SEQ ID NO: 31) | ATATGGTTTATGATGATGG (SEQ ID NO: 1043) |
| PCAF (SEQ ID NO: 33) | GAGCGGTAGGTGTCGAA (SEQ ID NO: 978) |
| PCAF (SEQ ID NO: 33) | GAGTGGTAGGTGTTGAA (SEQ ID NO: 979) |
| PCAF (SEQ ID NO: 33) | TAAGATTTTCGCGGGTA (SEQ ID NO: 980) |
| PCAF (SEQ ID NO: 33) | TGTAAGATTTTGTGGGTA (SEQ ID NO: 981) |
| PCAF (SEQ ID NO: 33) | AGTTCGTAGTTTCGAG (SEQ ID NO: 982) |
| PCAF (SEQ ID NO: 33) | GTTTGTAGTTTTGAGGA (SEQ ID NO: 983) |
| PCAF (SEQ ID NO: 33) | TAGGGCGCGGAGTAGA (SEQ ID NO: 984) |
| PCAF (SEQ ID NO: 33) | TAGGGTGTGGAGTAGA (SEQ ID NO: 985) |
| PRKCD (SEQ ID NO: 34) | ATTTATTTTTTCGTTGTAGG (SEQ ID NO: 772) |
| PRKCD (SEQ ID NO: 34) | TATTTATTTTTTGTGTAGG (SEQ ID NO: 773) |
| PRKCD (SEQ ID NO: 34) | TTTCGGAAACGGGAAT (SEQ ID NO: 774) |
| PRKCD (SEQ ID NO: 34) | TAGTTTTGGAAATGGGA (SEQ ID NO: 775) |
| PRKCD (SEQ ID NO: 34) | GGACGGAGTTATCGGT (SEQ ID NO: 776) |
| PRKCD (SEQ ID NO: 34) | GGATGGAGTTATTGGTA (SEQ ID NO: 777) |
| PRKCD (SEQ ID NO: 34) | GTTTAGCGGAGGGATA (SEQ ID NO: 778) |
| PRKCD (SEQ ID NO: 34) | TGTTTAGTGGAGGGAT (SEQ ID NO: 779) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTGTTACGGTTTGAGAG (SEQ ID NO: 780) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTGTTATGGTTTGAGAGT (SEQ ID NO: 781) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTTGTTATAGTTTGAGAGT (SEQ ID NO: 782) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTTGTTACGGTTTGAG (SEQ ID NO: 783) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTTGTTATGGTTTGAGA (SEQ ID NO: 784) |
| ESR1 (exon8) (SEQ ID NO: 61) | TTTGTTATAGTTTGAGAG (SEQ ID NO: 785) |

Table 4: Numbers of censored and relapsed patients in randomly selected sample set of ER+, N0, untreated population.

| | Frequency | Percentage |
|----------------------|------------|--------------|
| Censored | 276 | 66.5 |
| Distant metastasis | 66 | 15.9 |
| Locoregional relapse | 49 | 11.8 |
| Contralateral breast | 24 | 5.8 |
| Sum | 415 | 100.0 |

Table 5: Numbers of censored and relapsed patients in ER+, N0, TAM treated population.

| | Frequency | Percentage |
|----------------------|------------|--------------|
| Censored | 485 | 89.6 |
| Distant metastasis | 31 | 5.7 |
| Locoregional relapse | 20 | 3.7 |
| Contralateral breast | 5 | 0.9 |
| Sum | 541 | 100.0 |

Table 6: Primers and Amplificates according to Example 3

| Forward primer SEQ ID NO: | Reverse primer SEQ ID NO: | Amplificate SEQ ID NO: | Amplificate number |
|------------------------------|------------------------------|---------------------------|--------------------|
| 1150 | 1151 | 1152 | 1 |
| 1153 | 1154 | 1155 | 2 |
| 1156 | 1157 | 1158 | 3 |
| 1159 | 1160 | 1161 | 4 |
| 1162 | 1163 | 1164 | 5 |
| 1165 | 1166 | 1167 | 6 |
| 1168 | 1169 | 1170 | 7 |
| 1171 | 1172 | 1173 | 8 |
| 1174 | 1175 | 1176 | 9 |
| 1177 | 1178 | 1179 | 10 |
| 1180 | 1181 | 1182 | 11 |